Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing

B. Adcock
Purdue Univ.

A. Hansen
Univ. of Cambridge

C. Poon
Univ. of Cambridge

B. Roman
Univ. of Cambridge

1 Introduction

In this paper we bridge the substantial gap between existing compressed sensing theory and its current use in real-world applications.[1]

We do so by introducing a new mathematical framework for overcoming the so-called coherence barrier. Our framework generalizes the three traditional pillars of compressed sensing—namely, *sparsity*, *incoherence* and *uniform random subsampling*—to three new concepts: *asymptotic sparsity*, *asymptotic incoherence* and *multilevel random subsampling*. As we explain, asymptotic sparsity and asymptotic incoherence are more representative of real-world problems—e.g. imaging—than the usual assumptions of sparsity and incoherence. For instance, problems in Magnetic Resonance Imaging (MRI) are both asymptotically sparse and asymptotically incoherent, and hence amenable to our framework.

The second important contribution of the paper is an analysis of a novel and intriguing effect that occurs in asymptotically sparse and asymptotically incoherent problems. Namely, *the success of compressed sensing is resolution dependent.*

As suggested by their names, asymptotic incoherence and asymptotic sparsity are only truly witnessed for reasonably large problem sizes. When the problem size is small, there is little to be gained from compressed sensing over classical linear reconstruction techniques. However, as we show in this paper, once the resolution of the problem is sufficiently large, compressed sensing can and will offer a substantial advantage. This is so-called resolution dependence.

This phenomenon has the following two important consequences, which are also summarized in Figure 1:

(i) Suppose one considers a compressed sensing experiment where the sampling device, the object to be recovered, the sampling strategy and subsampling percentage are all fixed, but the resolution is allowed to vary. Resolution dependence means that a compressed sensing reconstruction done at high resolutions (e.g. 2048×2048) will yield much higher quality when compared to full sampling than one done at a low resolution (e.g. 256×256). This phenomenon has an important consequence for practitioners investigating the usefulness of compressed sensing algorithms. A scientist carrying out an experiment at low resolution may well conclude that compressed sensing imparts limited benefits. However, a markedly different conclusion would be reached if the same experiment were to be performed at higher resolution.

(ii) Suppose we conduct a similar experiment, but we now use the same total number of samples N (instead of the same percentage) at low resolution as we take at high resolution. Intriguingly, the above result still holds: namely, the higher resolution reconstruction will yield substantially better results. This is true because the multilevel random sampling strategy successfully exploits asymptotic sparsity and asymptotic incoherence. Thus, with the same amount of total effort, i.e. the number of measurements, compressed sensing with multilevel sampling works as a *resolution enhancer*: it allows one to recover the fine details of an image in a way that is not possible with the lower resolution reconstruction.

On a broader note, resolution dependence and its consequences suggest the following advisory for practitioners: it is critical that simulations with compressed sensing be carried out with a

[1]This paper is part of a larger project on subsampling in applications. Further details, as well as codes and numerical examples, can be found on the project website `http://subsample.org`

Results of *asymptotic sparsity* and *asymptotic incoherence*

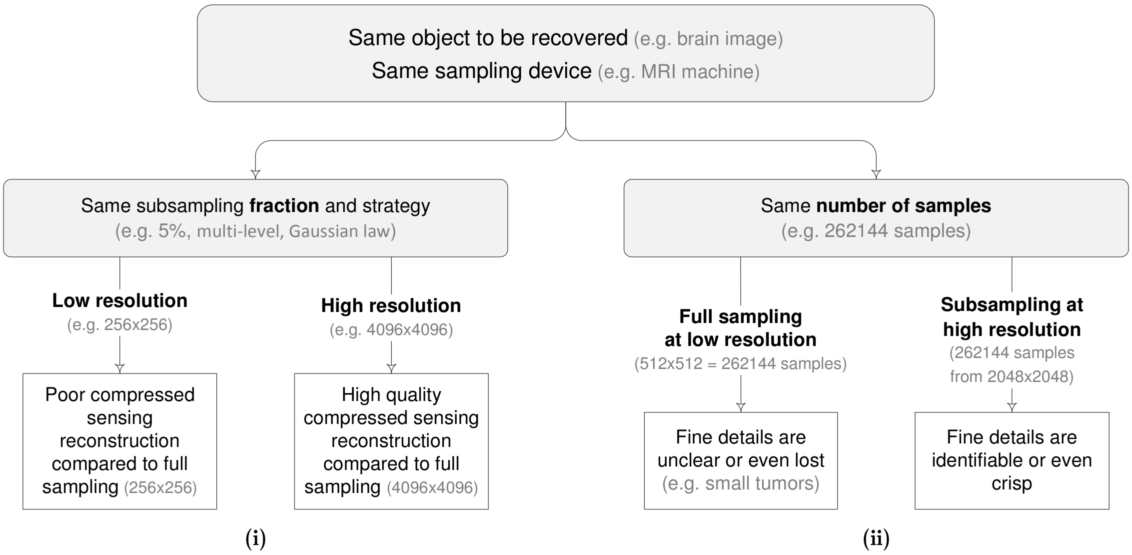


Figure 1: This illustrates one of the main messages of the paper: the success of compressed sensing is resolution dependent. (i) demonstrates how two identical experiments may give wildly different outcomes when performed with the same subsampling percentage at different resolution levels. (ii) demonstrates how the same amount of samples can give dramatically different outcomes at different resolution levels. In particular compressed sensing serves as a resolution enhancer.

careful understanding of the influence of the problem resolution. Naïve simulations with standard, low-resolution test images may very well lead to incorrect conclusions about the efficacy of compressed sensing as a tool for image reconstruction.

An important application of our work is the problem of MRI, which turns out to be a highly coherent problem. MRI served as one of the original motivations for compressed sensing, and continues to be a topic of substantial research. Some of the earliest work on this problem—in particular, the research of Lustig et al. [33, 34, 35]—demonstrated that, due to the high coherence, the standard random sampling strategies of compressed sensing theory lead to highly substandard reconstructions. On the other hand, random sampling according to some nonuniform density was shown empirically to lead to substantially improved reconstruction quality. Since the work of Lustig et al. these observations have been confirmed in numerous other investigations [34, 35, 38, 39, 45], and it is now standard in MR applications to use some sort of variable density strategy to overcome the coherence barrier.

This work has culminated in the extremely successful application of compressed sensing to MRI. However, a mathematical theory addressing these sampling strategies is largely lacking. Despite some recent work [31] (see Section 8 for a discussion), a substantial gap exists between the standard theorems of compressed sensing and its implementation in such problems.

The purpose of this paper is to introduce a mathematical foundation for compressed sensing for coherent problems and to rigorously show that the coherence barrier can be broken. In doing so, we provide a firm theoretical basis for the above empirical studies demonstrating the success of nonuniform density sampling. In addition, our main results give insight into how to design efficient subsampling techniques based on multi-level strategies.

Whilst the MR problem will serve as our main application, we stress that our theory is extremely general in that it holds for almost arbitrary sampling and sparsity systems. As we demonstrate, standard compressed sensing results, such as those of Candès, Romberg & Tao [14], Candès & Plan [12], are specific instances of our main theorems.

Another facet to our work is that we shall present theorems that cover not only the case of signals and images modelled as vectors in finite-dimensional vector spaces, but also elements of

separable Hilbert spaces. This continues the work of Adcock & Hansen [2] on *infinite-dimensional* compressed sensing. In Section 2.2 we explain the importance of this generalisation.

2 Background

2.1 Compressed sensing

Compressed sensing, introduced by Candès, Romberg & Tao [14] and Donoho [21], has been one of the major developments in applied mathematics in the last decade [10, 19, 25, 26, 27]. By exploiting a particular signal structure, namely sparsity, it allows one to circumvent the traditional barriers of sampling theory (e.g. the Nyquist rate) and recover signals and images from far fewer measurements than was classically considered possible. The list of applications of compressed sensing is diverse and growing, and includes medical imaging, analog-to-digital conversion, radar, and astronomy.

A typical setup in compressed sensing, and that which we shall in part follow in this paper, is as follows. Let $\{\psi_j\}_{j=1}^N$ and $\{\varphi_j\}_{j=1}^N$ be two orthonormal bases of $\mathbb{C}^N$, the *sampling* and *sparsity* bases respectively, and write

$$U = (u_{ij})_{i,j=1}^N \in \mathbb{C}^{N\times N}, \qquad u_{ij} = \langle \varphi_j, \psi_i \rangle.$$

Note that U is an isometry.

Definition 2.1. *Let $U = (u_{ij})_{i,j=1}^N \in \mathbb{C}^{N\times N}$ be an isometry. The coherence of U is precisely*

$$\mu(U) = \max_{i,j=1,\dots,N} |u_{ij}|^2 \in [N^{-1}, 1]. \tag{2.1}$$

We say that U is perfectly incoherent if $\mu(U) = N^{-1}$.

A signal $f \in \mathbb{C}^N$ is said to be s-sparse in the orthonormal basis $\{\varphi_j\}_{j=1}^N$ if at most s of its coefficients in this basis are nonzero. In other words, $f = \sum_{j=1}^N x_j \varphi_j$, and the vector $x \in \mathbb{C}^N$ satisfies $|\mathrm{supp}(x)| \leq s$, where

$$\mathrm{supp}(x) = \{j : x_j \neq 0\}.$$

Let $f \in \mathbb{C}^N$ be s-sparse in $\{\varphi_j\}_{j=1}^N$, and suppose we have access to the samples

$$\hat{f}_j = \langle f, \psi_j \rangle, \quad j = 1, \dots, N.$$

Let $\Omega \subseteq \{1, \dots, N\}$ be of cardinality m and chosen uniformly at random. According to a result of Candès & Plan [12] and Adcock & Hansen [2], f can be recovered exactly with probability exceeding $1 - \epsilon$ from the subset of measurements $\{\hat{f}_j : j \in \Omega\}$, provided

$$m \gtrsim \mu(U) \cdot N \cdot s \cdot \big(1 + \log(\epsilon^{-1})\big) \cdot \log N, \tag{2.2}$$

(here and elsewhere in this paper we shall use the notation $a \gtrsim b$ to mean that there exists a constant $C > 0$ independent of all relevant parameters such that $a \geq Cb$). In practice, recovery is achieved by solving the following convex optimization problem:

$$\min_{\eta \in \mathbb{C}^N} \|\eta\|_{l^1} \text{ subject to } P_\Omega U \eta = P_\Omega \hat{f}, \tag{2.3}$$

where $\hat{f} = (\hat{f}_1, \dots, \hat{f}_N)^\top$ and $P_\Omega \in \mathbb{C}^{N\times N}$ is the diagonal projection matrix with j^{th} entry 1 if $j \in \Omega$ and zero otherwise.

The key estimate (2.2) shows that the number of measurements m required is, up to a log factor, on the order of the sparsity s, provided the coherence $\mu(U) = \mathcal{O}\left(N^{-1}\right)$. This is the case, for example, when U is the DFT matrix; a problem which was studied in some of the first papers on compressed sensing [14] (this example is actually perfectly incoherent).

2.2 Continuous/infinite-dimensional problems

The framework of the previous section is suitable for many problems. However, there are some important problems where this framework can lead to significant errors, since the underlying problem is continuous, and hence not well represented by a finite-dimensional, vector space model [2, 15, 41]. To address this issue, a theory of compressed sensing in infinite dimensions was introduced by Adcock & Hansen in [2], based on a new approach to classical sampling known as *generalized sampling* [3, 4, 6, 5]. For implementation of a continuous/infinite-dimensional model of MRI using l^1 optimisation see [29].

Let us now describe the framework of [2] in more detail. Suppose that $\mathcal{H}$ is a separable Hilbert space over $\mathbb{C}$, and let $\{\psi_j\}_{j\in\mathbb{N}}$ be an orthonormal basis on $\mathcal{H}$ (the sampling basis). Let $\{\varphi_j\}_{j\in\mathbb{N}}$ be an orthonormal system in $\mathcal{H}$ (the sparsity system), and suppose that

$$U = (u_{ij})_{i,j\in\mathbb{N}}, \qquad u_{ij} = \langle \varphi_j, \psi_i \rangle, \tag{2.4}$$

is an infinite matrix. We may consider U as an element of $\mathcal{B}(l^2(\mathbb{N}))$; the space of bounded operators on $l^2(\mathbb{N})$ (throughout this paper we will make no distinction between bounded operators on sequence spaces and infinite matrices). As in the finite-dimensional case, U is an isometry, and we may define its coherence $\mu(U) \in (0,1]$ analogously to (2.1). Note, however, that $\mu(U)$ can be arbitrarily small in infinite dimensions.

We say that an element $f \in \mathcal{H}$ is (s,M)-sparse with respect to $\{\varphi_j\}_{j\in\mathbb{N}}$, where $s, M \in \mathbb{N}$, $s \leq M$, if the following holds:

$$f = \sum_{j\in\mathbb{N}} x_j \varphi_j, \qquad \mathrm{supp}(x) = \{j : x_j \neq 0\} \subseteq \{1,\ldots,M\}, \ |\mathrm{supp}(x)| \leq s.$$

Setting

$$\Sigma_{s,M} = \left\{ x \in l^2(\mathbb{N}) : x \text{ is } (s,M)\text{-sparse} \right\},$$

and

$$\sigma_{s,M}(f) = \min_{\eta\in\Sigma_{s,M}} \|x-\eta\|_{l^1}, \qquad f = \sum_{j\in\mathbb{N}} x_j\varphi_j, \quad x = (x_j)_{j\in\mathbb{N}} \in l^1(\mathbb{N}),$$

we say that f is (s,M)-compressible with respect to $\{\varphi_j\}_{j\in\mathbb{N}}$ if $\sigma_{s,M}(f)$ is small. Whenever f is (s,M)-sparse or compressible, we seek to recover it from a small number of the measurements

$$\hat{f}_j = \langle f, \psi_j \rangle, \quad j \in \mathbb{N}.$$

To do this, we introduce a second parameter $N \in \mathbb{N}$, and let Ω be a randomly chosen subset of indices $1,\ldots,N$ of size m. In the absence of noise we now solve

$$\inf_{\eta\in l^1(\mathbb{N})} \|\eta\|_{l^1} \text{ subject to } P_\Omega U P_M \eta = P_\Omega \hat{f}, \tag{2.5}$$

where $\hat{f} = (\hat{f}_j)_{j\in\mathbb{N}} \in l^2(\mathbb{N})$ and P_Ω is the projection operator corresponding to the index set Ω. In [2] it was proved that any solution to (2.5) recovers f exactly up to an error determined by $\sigma_{k,M}(f)$, provided N and m satisfy an appropriate *balancing property* with respect to M and s (see Definition 4.3), and provided

$$m \gtrsim \mu(U) \cdot N \cdot s \cdot \left(1 + \log(\epsilon^{-1})\right) \cdot \log\left(m^{-1} M N \sqrt{s}\right).$$

As in the finite-dimensional case (which is a corollary of the result), we find that m is on the order of the sparsity s whenever $\mu(U)$ is sufficiently small.

We shall discuss this result, and in particular, the nature of the balancing property, in more detail in §4.1.2. However, we remark in passing that it is usually not sufficient to take $N = M$, and doing so can quite easily lead to substantial errors [2]. In general, one requires $N > M$ for the balancing property to hold.

Note that this framework generalizes finite-dimensional compressed sensing theory in a natural way—vector spaces are replaced by separable Hilbert spaces—and known results, such as that described in Section 2.1, are straightforward corollaries of theorems proved in [2]. We shall proceed in a similar manner in this paper: finite-dimensional theorems will be corollaries of those pertaining to the infinite-dimensional case.

2.3 The coherence barrier

In either finite- or infinite-dimensional compressed sensing, the number of measurements required is, up to a log factor, on the order of the sparsity s multiplied by $\mu(U)N$. When the coherence (or mutual coherence, as it is often known [22, 23]) is small, the energy of the signal is sufficiently spread out amongst its samples to allow for recovery using only $\mathcal{O}(s)$ measurements. On the other hand, when $\mu(U)$ large, one cannot expect to reconstruct an s-sparse vector f from highly subsampled measurements, regardless of the recovery algorithm employed [12]. We refer to this as the *coherence barrier*.

The MRI problem is an important instance of this barrier:

Example 2.1 (Finite-dimensional MRI model) Suppose that $f \in \mathbb{C}^N$ is sparse in a discrete wavelet basis $\{\varphi_j\}_{j=1}^N$, and let $\{\psi_j\}_{j=1}^N$ be the rows of $N \times N$ discrete Fourier transform (DFT) matrix. In this case, the matrix $U = \mathrm{DFT} \cdot \mathrm{DWT}^{-1}$ satisfies $\mu(U) = \mathcal{O}(1)$ for any N [13, 31].

Example 2.2 (Infinite-dimensional MRI model) Let $f \in L^2(\mathbb{R})$ have compact support, and suppose that f is sparse in an orthonormal system of compactly supported wavelets $\{\varphi_j\}_{j\in\mathbb{N}}$. Let $\{\psi_j\}_{j\in\mathbb{N}}$ be the standard Fourier basis, i.e. $\psi_j(x) = \sqrt{\epsilon}\mathrm{e}^{2\pi \mathrm{i} j \epsilon x}$, $j \in \mathbb{Z}$ (note that we enumerate over $\mathbb{Z}$ as opposed to $\mathbb{N}$ in this case) for suitable $\epsilon > 0$ (see Section 6.4 for details on this construction). In this case, $\mu(U) = \mathcal{O}(1)$ for any such wavelet basis (Theorem 3.2). For a continuous/infinite-dimensional model of MRI using l^1 optimisation see [29].

These two problems will serve as our main examples throughout the paper.

3 New concepts

We now discuss the main concepts of the paper: namely, asymptotic incoherence, asymptotic sparsity and multilevel sampling. We shall work primarily in the infinite-dimensional setting of §2.2, with the finite-dimensional case being a straightforward corollary.

3.1 Asymptotic incoherence

Consider Example 2.2 in the case where $\mathcal{H} = L^2(0,1)$, $\epsilon = 1$ and $\{\varphi_j\}_{j\in\mathbb{N}}$ is the Haar wavelet basis. Note that the coherence $\mu(U)$ for this problem is exactly one. In Figure 2 we plot the absolute values of the entries of the matrix U. As is evident, the larger values of U are located near its centre (recall that we enumerate over $\mathbb{Z}$ for the sampling basis and $\mathbb{N}$ for the sparsity basis), and as one moves away from this region the values get progressively smaller.

This motivates the following general definition:

Definition 3.1. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Then U is asymptotically incoherent if*

$$\mu(P_N^\perp U),\ \mu(U P_N^\perp) \to 0, \quad N \to \infty. \tag{3.1}$$

Here $P_N \in \mathcal{B}(l^2(\mathbb{N}))$ is the projection operator onto $\mathrm{span}\{e_j : j = 1, \dots, N\}$, where $\{e_j\}_{j\in\mathbb{N}}$ is the canoncial basis for $l^2(\mathbb{N})$. In other words, U is asymptotically incoherent if the coherence of the infinite matrices formed by replacing either the first N rows or columns of U by zeros tends to zero as $N \to \infty$.

Asymptotic incoherence in the case of Haar wavelets with Fourier sampling is indicated by Figure 2. As it happens, this is always the case for the problem of Example 2.2, regardless of the wavelet basis used. We have

Theorem 3.2. *Consider the setup of Example 2.2. Then $\mu(U) \geq \epsilon |\hat{\Phi}(0)|^2$, where Φ is the corresponding scaling function, and $\mu(P_N^\perp U), \mu(U P_N^\perp) = \mathcal{O}\left(N^{-1}\right)$ as $N \to \infty$.*

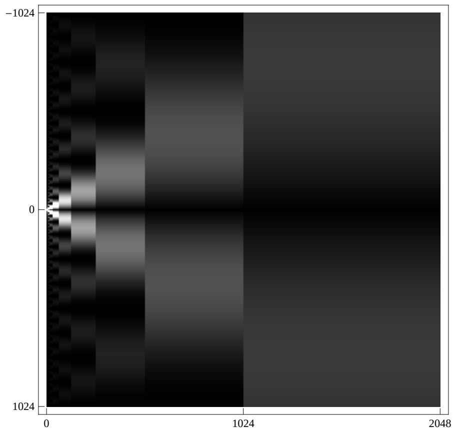


Figure 2: The absolute values of the matrix U for Haar wavelets with Fourier sampling. Light regions correspond to large values and dark regions to small values. Observe the asymptotic incoherence as described by Theorem 3.2.

3.2 Multi-level sampling

Asymptotic incoherence suggests a different subsampling strategy should be used instead of standard random sampling. High coherence in the first few rows of U means that important information about the signal to be recovered may well be contained in its corresponding measurements. Hence to ensure good recovery we should fully sample these rows. Once outside of this region, when the coherence starts to decrease, we can begin to subsample. Let $N_1, N, m \in \mathbb{N}$ be given. This now leads us to consider an index set Ω of the form $\Omega = \Omega_1 \cup \Omega_2$, where $\Omega_1 = \{1, \ldots, N_1\}$, and $\Omega_2 \subseteq \{N_1 + 1, \ldots, N\}$ is chosen uniformly at random with $|\Omega_2| = m$. We refer to this as a *two-level* sampling scheme. As we shall prove later, the amount of subsampling possible (i.e. the parameter m) in the region corresponding to Ω_2 will depend solely on the sparsity of the signal and coherence $\mu(P_{N_1}^{\perp} U)$.

The two-level scheme represents the simplest type of nonuniform density sampling. There is no reason, however, to restrict our attention to just two levels (full and subsampled). In general, we shall consider *multilevel* schemes, defined as follows:

Definition 3.3. *Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \ldots < N_r$, $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$, with $m_k \leq N_k - N_{k-1}$, $k = 1, \ldots, r$, and suppose that*

$$\Omega_k \subseteq \{N_{k-1} + 1, \ldots, N_k\}, \quad |\Omega_k| = m_k, \quad k = 1, \ldots, r,$$

are chosen uniformly at random, where $N_0 = 0$. We refer to the set

$$\Omega = \Omega_{\mathbf{N},\mathbf{m}} := \Omega_1 \cup \ldots \cup \Omega_r.$$

as an $(\mathbf{N}, \mathbf{m})$-multilevel sampling scheme.

The same guiding principle applies as in the two-level case. In the region of highest coherence, i.e. Ω_1, we take more measurements, and as coherences decreases, i.e. as the level number k increases, we take progressively fewer. Note that our introduction of multilevel schemes is not just for the purposes of mathematical intricacy: in practice, they are often more effective than two-level schemes.

Note that similar sampling strategies are found in most empirical studies on compressive MRI [34, 35, 38, 39]. A closely related strategy was considered by Candès & Romberg in [13] (see also [45]), where the sampling levels correspond precisely to the wavelet scales. Our theory generalizes this approach by removing this condition (see Remark 3.1). Moreover, our theorems also do not require the image to be first separated into individual subbands before sampling, such as the case in [13].

Another instance of a two-level strategy is found in [42]. Here the authors consider application of compressed sensing in fluorescence microscopy via a so-called "half-half" scheme.

3.3 Asymptotic sparsity in levels

In the case of perfect incoherence, the standard random sampling strategies of compressed sensing are highly effective for sparse signals. However, in asymptotically incoherent setting, the notion of sparsity can be substantially relaxed.

To explain this, let $x = (x_j)_{j\in\mathbb{N}} \in l^1(\mathbb{N})$ be the infinite vector of coefficients of a function f in the sparsity system $\{\varphi_j\}_{j\in\mathbb{N}}$. Suppose that x was very sparse in its entries $j = 1, \ldots, M_1$, but the exact location of these nonzero coefficients x_j was unknown. Since the matrix U is highly coherent in its corresponding rows, there is no way we can exploit this sparsity to achieve subsampling. High coherence forces us to sample fully the first M_1 rows, otherwise we run the risking of missing critical information about x. On the other hand, once the asymptotic incoherence sets in we are able to subsample the rows of U and recover sparse sets of coefficients.

This means that there is nothing to be gained from high sparsity of x in its first few entries. However, we can expect to achieve significant subsampling if x is *asymptotically sparse*, i.e. the coefficients x_j, $j > M_1$, are sparse for some sufficiently large M_1. This motivates the following two definitions:

Definition 3.4. *For $r \in \mathbb{N}$ let $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ with $1 \leq M_1 < \ldots < M_r$ and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, with $s_k \leq M_k - M_{k-1}$, $k = 1, \ldots, r$, where $M_0 = 0$. We say that $x \in l^2(\mathbb{N})$ is $(\mathbf{s}, \mathbf{M})$-sparse if, for each $k = 1, \ldots, r$,*

$$\Delta_k := \mathrm{supp}(x) \cap \{M_{k-1} + 1, \ldots, M_k\},$$

satisfies $|\Delta_k| \leq s_k$. We denote the set of $(\mathbf{s}, \mathbf{M})$-sparse vectors by $\Sigma_{\mathbf{s},\mathbf{M}}$.

Definition 3.5. *Let $f = \sum_{j\in\mathbb{N}} x_j \varphi_j \in \mathcal{H}$, where $x = (x_j)_{j\in\mathbb{N}} \in l^1(\mathbb{N})$. We say that f is $(\mathbf{s}, \mathbf{M})$-compressible with respect to $\{\varphi_j\}_{j\in\mathbb{N}}$ if $\sigma_{\mathbf{s},\mathbf{M}}(f)$ is small, where*

$$\sigma_{\mathbf{s},\mathbf{M}}(f) := \min_{\eta\in\Sigma_{\mathbf{s},\mathbf{M}}} \|x - \eta\|_{l^1}. \tag{3.2}$$

As we shall explain, signals possessing this sparsity pattern—which we henceforth refer to as being *asymptotically sparse in levels*—are ideally suited to multilevel sampling schemes. Roughly speaking, the number of measurements m_k required in each band Ω_k is determined by the sparsity of f in the corresponding band Δ_k and the asymptotic coherence.

Remark 3.1 In Section 5.1 we shall show that natural images are asymptotically compressible when the levels $\mathbf{M}$ correspond to the wavelet scales. In this case it is somewhat natural, although not necessary, to employ a multilevel sampling scheme corresponding exactly to these levels. As mentioned in Section 3.2, this particular approach was previously considered in [13].

4 Main theorems

4.1 Two-level sampling schemes

4.1.1 The sparse and noiseless case

We shall commence with the finite-dimensional result concerning exact recovery, but before that we need a definition of local coherence:

Definition 4.1. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Given $N \in \mathbb{N}$ we define*

$$\mu_N = \mu(P_N^{\perp} U).$$

If $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ and $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \ldots N_r$ and $1 \leq M_1 < \ldots < M_r$ we define the $(k,l)^{\text{th}}$ local coherence of U with respect to $\mathbf{N}$ and $\mathbf{M}$ by

$$\mu_{\mathbf{N},\mathbf{M}}(k,l) = \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}}) \cdot \mu(P_{N_{k-1}}^{\perp} U)}, \quad k, l = 1, \ldots, r,$$

where $N_0 = M_0 = 0$ and P_b^a denotes the projection corresponding to indices $\{a + 1, \ldots, b\}$.

We can now state the first main theorem.

Theorem 4.2. *Let $U \in \mathbb{C}^{N\times N}$ be an isometry and $x \in \mathbb{C}^N$ be $(\mathbf{s},\mathbf{M})$-sparse, where $r = 2$, $\mathbf{s} = (M_1, s_2)$, $s = M_1 + s_2$ and $\mathbf{M} = (M_1, M_2)$ with $M_2 = N$. Suppose that*

$$\|P_{N_1}^{\perp} U P_{M_1}\| \leq \frac{\gamma}{\sqrt{M_1}}, \tag{4.1}$$

for some $1 \leq N_1 \leq N$ and $\gamma \in (0, 2/5]$, and that $\gamma \leq s_2\sqrt{\mu_{N_1}}$. For $\epsilon > 0$, let $m \in \mathbb{N}$ satisfy

$$m \gtrsim (N - N_1) \cdot (\log(s\epsilon^{-1}) + 1) \cdot \mu_{N_1} \cdot s_2 \cdot \log(N).$$

Let $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ be a two-level sampling scheme, where $\mathbf{N} = (N_1, N_2)$ and $\mathbf{m} = (m_1, m_2)$ with $N_2 = N$, $m_1 = N_1$ and $m_2 = m$, and suppose that $\xi \in \mathbb{C}^N$ is a minimizer of (2.3), where $\hat{f} = Ux$. Then, with probability exceeding $1 - \epsilon$, ξ is unique and $\xi = x$.

Note that the proof of this theorem includes the standard random sampling compressed sensing results of §2.1 as a special case. This follows by allowing $M_1 = N_1 = 0$, in which case (4.1) is redundant.

To state the corresponding result in the infinite-dimensional case, we require the following definition of the balancing property [2]:

Definition 4.3. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Then $N \in \mathbb{N}$ and $K \geq 1$ satisfy the weak balancing property with respect to U, $M \in \mathbb{N}$ and $s \in \mathbb{N}$ if*

$$\|P_M U^* P_N U P_M - P_M\|_{l^\infty \to l^\infty} \leq \frac{1}{8}\left(\log_2^{1/2}\left(4\sqrt{s}KM\right)\right)^{-1}, \tag{4.2}$$

where $\|\cdot\|_{l^\infty\to l^\infty}$ is the norm on $\mathcal{B}(l^\infty(\mathbb{N}))$. We say that N and K satisfy the strong balancing property with respect to U, M and s if (4.2) holds, as well as

$$\|P_M^{\perp} U^* P_N U P_M\|_{l^\infty\to l^\infty} \leq \frac{1}{8}. \tag{4.3}$$

We now have the following:

Theorem 4.4. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry and $x \in l^1(\mathbb{N})$ be $(\mathbf{s},\mathbf{M})$-sparse, where $r = 2$, $\mathbf{s} = (M_1, s_2)$ and $\mathbf{M} = (M_1, M_2)$. Suppose $N_1, N_2, m_2 \in \mathbb{N}$ are such that the parameters*

$$N := N_2, \quad K := (N_2 - N_1)/m_2,$$

satisfy the weak balancing property with respect to U, $M := M_2$ and $s := M_1 + s_2$, and that, for some $\gamma \in (0, 2/5]$,

$$\|P_{N_1}^{\perp} U P_{M_1}\| \leq \frac{\gamma}{\sqrt{M_1}},$$

and $\gamma \leq s_2\sqrt{\mu_{N_1}}$. For $\epsilon > 0$, let $m \in \mathbb{N}$ satisfy

$$m \gtrsim (N - N_1) \cdot (\log(s\epsilon^{-1}) + 1) \cdot \mu_{N_1} \cdot s_2 \cdot \log\left(KM\sqrt{s}\right), \tag{4.4}$$

and suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a two-level sampling scheme, where $\mathbf{N} = (N_1, N_2)$ and $\mathbf{m} = (m_1, m_2)$ with $m_1 = N_1$ and $m_2 = m$. Let $\xi \in l^1(\mathbb{N})$ be a minimizer of (2.5), where $\hat{f} = Ux$. Then, with probability exceeding $1 - \epsilon$, ξ is unique and coincides with x. If $m = N - N_1$ then this holds with probability 1.

Note that this theorem generalizes the infinite-dimensional compressed sensing result of [2] (see also §2.2) to the two-level sampling case.

4.1.2 The role of the balancing property

The main difference between the finite- and infinite-dimensional theorems (Theorems 4.2 and 4.4) is that the parameters N and K in the latter must satisfy the weak balancing property.

The balancing property ensures that the truncated matrix $P_N U P_M$ is close to an isometry. In reconstruction problems, the presence of an isometry ensures stability in the mapping between measurements and coefficients [3], which explains the need for a such a property in our theorems. As explained in [2], without the balancing property the lack of stability in the underlying mapping leads frequently to numerically useless reconstructions.

Note that the balancing property does always hold, provided N is chosen sufficiently large in comparison to M. For details we refer to [2]. On the other hand, no balancing property is required in the finite-dimensional case since $P_N U P_M \equiv U$ is an isometry by assumption.

4.1.3 The noisy, nonsparse case

In realistic problems, signals are never exactly sparse (or asymptotically sparse), and their measurements are always contaminated by noise. Let $f = \sum_j x_j \varphi_j$ be a fixed signal, and write

$$y = P_\Omega \hat{f} + z = P_\Omega U x + z,$$

for its noisy measurements, where $z \in \mathrm{ran}(P_\Omega)$ is a noise vector satisfying $\|z\| \leq \delta$ for some $\delta \geq 0$. If δ is known, we now consider the following problem:

$$\inf_{\eta \in \mathcal{H}} \|\eta\|_{l^1} \text{ subject to } \|P_\Omega U \eta - y\| \leq \delta. \tag{4.5}$$

We now state our main result on (4.5) for two-level schemes. Since the finite-dimensional case is a straightforward corollary of the infinite-dimensional result, we present only the latter:

Theorem 4.5. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry and $x \in l^1(\mathbb{N})$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a two-level sampling scheme, where $\mathbf{N} = (N_1, N_2)$ and $\mathbf{m} = (N_1, m_2)$. Let $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, M_2) \in \mathbb{N}^2$, $M_1 < M_2$, and $\mathbf{s} = (M_1, s_2) \in \mathbb{N}^2$, be any pair such that the following holds:*

(i) we have $\|P_{N_1}^{\perp} U P_{M_1}\| \leq \frac{\gamma}{\sqrt{M_1}}$ and $\gamma \leq s_2 \sqrt{\mu_{N_1}}$ for some $\gamma \in (0, 2/5]$;

(ii) the parameters

$$N := N_2, \qquad K := (N_2 - N_1)/m_2$$

satisfy the strong balancing property with respect to U, $M := M_2$ and $s := M_1 + s_2$;

(iii) for $\epsilon > 0$, let

$$m_2 \gtrsim (N - N_1) \cdot (\log(s\epsilon^{-1}) + 1) \cdot \mu_{N_1} \cdot s_2 \cdot \log\left(KM\sqrt{s}\right).$$

Suppose that $\xi \in l^1(\mathbb{N})$ is a minimizer of (4.5). Then, with probability exceeding $1 - \epsilon$, we have

$$\|\xi - x\| \leq C\left(1 + \sqrt{K}\right) \cdot \left(\delta \cdot \left(1 + L \cdot \sqrt{s}\right) + \sigma_{\mathbf{s},\mathbf{M}}(f)\right), \tag{4.6}$$

for some constant C, where $\sigma_{\mathbf{s},\mathbf{M}}(f)$ is as in (3.2), and $L = \sqrt{6K}\left(\frac{3}{2} + \sqrt{1 + \frac{\log_2(6\epsilon^{-1}) + 1}{\log_2(4KM\sqrt{s})}}\right)$. If $m_2 = N - N_1$ then this holds with probability 1.

Note that the constant C in (4.6) (and also in Theorems 4.7 and 4.8 later) is exactly the constant of Proposition 6.1, and can bound found explicitly by following the steps of the proof. However, we have made no attempts to optimize this constant, hence we leave it in this form.

4.1.4 Discussion

Theorems 4.2 and 4.4 demonstrate that asymptotic incoherence and two-level sampling overcomes the coherence barrier. To see this, note the following:

(i) The condition $\|P_{N_1}^{\perp} U P_{M_1}\| \leq \frac{2}{5\sqrt{M_1}}$ (which is always satisfied for some N_1, since U is an isometry) implies that fully sampling the first N_1 measurements allows one to recover the first M_1 coefficients of f.

(ii) To recover the remaining s_2 coefficients we require, up to log factors, an additional

$$m_2 \gtrsim (N - N_1) \cdot \mu_{N_1} \cdot s_2,$$

measurements, taken randomly from the range $M_1 + 1, \ldots, M_2$. In particular, if N_1 is a fixed fraction of N, and if $\mu_{N_1} = \mathcal{O}\left(N_1^{-1}\right)$, such as for wavelets with Fourier measurements (Theorem 3.2), then one requires only $m_2 \gtrsim s_2$ additional measurements to recover the sparse part of the signal.

(iii) When f is asymptotically sparse, such is the case for natural images (see Section 5.1), then the relative size of s_2 will become smaller as M and N grow. In particular, the percentage $\left(\frac{N_1+m_2}{N}\right) \times 100$ of measurements required will decrease. Hence the subsampling rate possible will improve as the problem resolution becomes larger (see Section 5.2).

Note also that the two-level sampling scheme is completely robust in the presence of noise and inexact asymptotic sparsity, as shown by Theorem 4.5.

Remark 4.1 It is not necessary to know the sparsity structure, i.e. the values $\mathbf{s}$ and $\mathbf{M}$, of the image f in order to implement the two-level sampling technique (the same also applies to the multilevel technique discussed in the next section). Given a two-level scheme $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$, Theorem 4.5 demonstrates that f will be recovered exactly up to an error on the order of $\sigma_{\mathbf{s},\mathbf{M}}(f)$, where $\mathbf{s}$ and $\mathbf{M}$ are determined implicitly by $\mathbf{N}$, $\mathbf{m}$ and the conditions (i)–(iii) of the theorem. Of course, some *a priori* knowledge of $\mathbf{s}$ and $\mathbf{M}$ will greatly assist in selecting the parameters $\mathbf{N}$ and $\mathbf{m}$ so as to get the best recovery results. However, this is not necessary for implementing the method.

Remark 4.2 To simplify their presentation, Theorems 4.2, 4.4 and 4.5 contain the additional condition $\gamma \leq s_2\sqrt{\mu_{N_1}}$. This is a lower bound for the sparsity s_2 in the second level. It is possible to remove this condition, in which case the corresponding estimate for m_2 will be

$$m_2 \gtrsim (N - N_1) \cdot \left(\gamma \cdot \sqrt{\mu_{N_1}} + \mu_{N_1} \cdot s_2\right), \tag{4.7}$$

plus log factors. Note that adding the constraint $\gamma \leq s_2\sqrt{\mu_{N_1}}$ merely reduces (4.7) to that given in the theorems. As we explain in Section 4.2.3, the reason for the additional term in (4.7) is due to the phenomenon of interference between the two sparsity levels. Fortunately, however, the condition $\gamma \leq s_2\sqrt{\mu_{N_1}}$ is always satisfied in practice. Typically in practice both s_2 and N_1 are fixed percentages of the total resolution N (see Section 5.2). If $\mu_{N_1} = \mathcal{O}(1/N_1)$, for example, then this condition will always hold for all reasonably large N.

4.2 Multilevel sampling schemes

We now consider multilevel sampling schemes with arbitrary numbers of levels. Before we present our main theorems we require the following definition:

Definition 4.6. *Let U be an isometry of either $\mathbb{C}^{N\times N}$ or $\mathcal{B}(l^2(\mathbb{N}))$. For $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$, $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \ldots < N_r$ and $1 \leq M_1 < \ldots < M_r$, $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$ and $1 \leq k \leq r$, let*

$$S_k = S_k(\mathbf{N}, \mathbf{M}, \mathbf{s}) = \max_{\eta \in \Theta} \|P_{N_k}^{N_{k-1}} U \eta\|^2,$$

where $N_0 = M_0 = 0$ and Θ is given by

$$\Theta = \{\eta : \|\eta\|_{l^\infty} \leq 1, |\mathrm{supp}(P_{M_l}^{M_{l-1}} \eta)| = s_l, \, l = 1, \ldots, r\}.$$

4.2.1 The finite-dimensional case

We start with the finite-dimensional case. For brevity, we now only present our results for the noisy, nonsparse case. The corresponding theorems for the exactly sparse, noise-free case are straightforward corollaries.

Theorem 4.7. *Let $U \in \mathbb{C}^{N\times N}$ be an isometry and $x \in \mathbb{C}^N$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme, where $\mathbf{N} = (N_1,\dots,N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1,\dots,m_r) \in \mathbb{N}^r$. Let $(\mathbf{s},\mathbf{M})$, where $\mathbf{M} = (M_1,\dots,M_r) \in \mathbb{N}^r$, $M_1 < \dots < M_r$, and $\mathbf{s} = (s_1,\dots,s_r) \in \mathbb{N}^r$, be any pair such that the following holds: for $\epsilon > 0$ and $1 \le k \le r$,*

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot (\log(s\epsilon^{-1}) + 1) \cdot (\mu_{\mathbf{N},\mathbf{M}}(k,1) \cdot s_1 + \dots + \mu_{\mathbf{N},\mathbf{M}}(k,r) \cdot s_r) \cdot \log(N), \tag{4.8}$$

where $s := s_1 + \dots + s_r$ and

$$m_k \gtrsim \hat{m}_k \cdot (\log(s\epsilon^{-1}) + 1) \cdot \log(N),$$

where $\hat{m}_k$ satisfies

$$1 \ge \left(\frac{N_1 - N_0}{\hat{m}_1} - 1\right) \cdot \mu_{N_0} \cdot \tilde{s}_1 + \dots + \left(\frac{N_r - N_{r-1}}{\hat{m}_r} - 1\right) \cdot \mu_{N_{r-1}} \cdot \tilde{s}_r, \tag{4.9}$$

for any $\tilde{s}_1,\dots,\tilde{s}_r \in (0,\infty)$ such that

$$\tilde{s}_1 + \dots + \tilde{s}_r \le s_1 + \dots + s_r, \qquad \tilde{s}_k \le S_k(\mathbf{N},\mathbf{M},\mathbf{s}).$$

Suppose that $\xi \in l^1(\mathbb{N})$ is a minimizer of (4.5). Then, with probability exceeding $1-\epsilon$, we have that

$$\|\xi - x\| \le C \cdot \left(1 + \sqrt{K}\right) \cdot \left(\delta \cdot \left(1 + L \cdot \sqrt{s}\right) + \sigma_{\mathbf{s},\mathbf{M}}(f)\right), \tag{4.10}$$

for some constant C, where $\sigma_{\mathbf{s},\mathbf{M}}(f)$ is as in (3.2), $L = \sqrt{6K}\left(\frac{3}{2} + \sqrt{1 + \frac{\log_2(\epsilon^{-1})+1}{\log_2(4KM\sqrt{s})}}\right)$ and $K = \max_{k=1,\dots,r}\left\{\frac{N_k - N_{k-1}}{m_k}\right\}$. If $m_k = N_k - N_{k-1}$,$1 \le k \le r$, then this holds with probability 1.

4.2.2 Example: the block-diagonal case

The most important part of Theorem 4.7 are the bounds for the number of samples m_k required in the k^{th} level of the multilevel scheme. This depends completely on the behaviours of the quantity $S_k(\mathbf{N},\mathbf{M},\mathbf{s})$ and the local coherences $\mu_{\mathbf{N},\mathbf{M}}(k,l)$.

Consider the case where the isometry $U \in \mathbb{C}^{N\times N}$ is a block diagonal matrix, with the k^{th} block being of size $(N_k - N_{k-1}) \times (M_k - M_{k-1})$. In this setting, $S_k = s_k$ for all k, and therefore we may set $\tilde{s}_k = s_k$ in Theorem 4.7. Since one also has in this case that $\mu_{\mathbf{N},\mathbf{M}}(k,l) = 0$ whenever $k \ne l$, and since $\mu(\mathbf{N},\mathbf{M})(k,k) \le \mu_{N_{k-1}}$, equations (4.8) and (4.9) reduce to

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot \left(\log(s\epsilon^{-1}) + 1\right) \cdot \mu_{N_{k-1}} \cdot s_k \cdot \log N,$$

and

$$1 \ge \left(\frac{N_1 - N_0}{\hat{m}_1} - 1\right) \cdot \mu_{N_0} \cdot s_1 + \dots + \left(\frac{N_r - N_{r-1}}{\hat{m}_r} - 1\right) \cdot \mu_{N_{r-1}} \cdot s_r.$$

In particular, it suffices to take

$$m_k \gtrsim (N_k - N_{k-1}) \cdot \left(\log(s\epsilon^{-1}) + 1\right) \cdot \mu_{N_{k-1}} \cdot s_k \cdot \log N, \quad 1 \le k \le r.$$

This is as one might expect: the number of measurements in the k^{th} level depends on the size of the level multiplied by the asymptotic incoherence and the sparsity in that level.

4.2.3 Sharpness of the estimates

When U is not block diagonal the inequalities in Theorem 4.7 cannot be reduced to such a simple form. In general, there can be interference between different sparsity levels, which means that S_k can be larger than s_k. Indeed, in general one has the upper bound

$$S_k \leq s = s_1 + \ldots + s_r,$$

and it is possible to show that there matrices U for which $S_k = s$ for all $k = 1, \ldots, r$. In such cases Theorem 4.7 predicts that m_k must scale at least like the total sparsity s, as opposed to the local sparsity s_k.

This actually turns out to be a sharp result. Consider the following construction. Let $N = rn$ for some $n \in \mathbb{N}$ and $\mathbf{N} = \mathbf{M} = (n, 2n, \ldots, rn)$. Let $W \in \mathbb{C}^{n \times n}$ be any isometry, and suppose that $V \in \mathbb{C}^{r \times r}$ is a full isometry, i.e. all its entries are nonzero (for example, V could be the DFT matrix). Let $U = V \otimes W$, where $\otimes$ is the usual Kronecker product, and note that $U \in \mathbb{C}^{N \times N}$ is also an isometry. Now suppose that $x = (x_1, \ldots, x_r) \in \mathbb{C}^N$ is an $(\mathbf{s}, \mathbf{M})$-sparse vector, where each $x_k \in \mathbb{C}^n$ is s_k-sparse. Suppose also that $s \leq n/r$ and that the x_k's have disjoint support sets, i.e. $\mathrm{supp}(x_k) \cap \mathrm{supp}(x_l) = \emptyset$, $k \neq l$. Then one has that

$$Ux = y, \quad y = (y_1, \ldots, y_r),\ y_k = W z_k,\ z_k = \sum_{l=1}^{r} v_{kl} x_l.$$

Hence the problem of recovering x from measurements y with an $(\mathbf{N}, \mathbf{m})$-multilevel strategy decouples into r problems of recovering the vector z_k from measurements $y_k = W z_k$, $k = 1, \ldots, r$. However, by construction each vector z_k is exactly s-sparse. Hence, since the coherence provides an information-theoretic limit [12], one requires at least $m_k \gtrsim n \cdot \mu(W) \cdot s \cdot \log n$ measurements at level k in order to recover each z_k, and therefore recover x, regardless of the reconstruction method used. This is exactly as predicted by Theorem 4.7 for this example.

Fortunately, such cases are extreme, and do not arise in the main application considered in this paper, namely that of recovering wavelet coefficients from Fourier samples. In fact, in this case, numerical results suggest that $S_k \approx s_k$, and therefore the subsampling permitted at level k is dictated mostly by the sparsity s_k and the asymptotic coherence $\mu_{N_{k-1}}$. Note that the other local coherences $\mu_{\mathbf{N},\mathbf{M}}(k, l)$ with $l \neq k$ do have an effect. However, this tends to be negligible, since they are usually much smaller than $\mu_{\mathbf{N},\mathbf{M}}(k, k)$ in practice.

4.2.4 The infinite-dimensional case

In the infinite-dimensional setting our main theorem is as follows:

Theorem 4.8. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry and $x \in l^1(\mathbb{N})$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme, where $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$. Let $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$, $M_1 < \ldots < M_r$, and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, be any pair such that the following holds:*

(i) the parameters

$$N := N_r, \quad K := \max_{k=1,\ldots,r} \left\{ \frac{N_k - N_{k-1}}{m_k} \right\},$$

satisfy the strong balancing property with respect to U, $M := M_r$ and $s := s_1 + \ldots + s_r$;

(ii) for $\epsilon > 0$ and $1 \leq k \leq r$,

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot (\log(s\epsilon^{-1}) + 1) \cdot (\mu_{\mathbf{N},\mathbf{M}}(k, 1) \cdot s_1 + \ldots + \mu_{\mathbf{N},\mathbf{M}}(k, r) \cdot s_r) \cdot \log\left(K \tilde{M} \sqrt{s}\right),$$

where $s := s_1 + \ldots + s_r$ and

$$m_k \gtrsim \hat{m}_k \cdot (\log(s\epsilon^{-1}) + 1) \cdot \log\left(K \tilde{M} \sqrt{s}\right),$$

where $\hat{m}_k$ satisfies (4.9) and

$$\tilde{M} = \min\{i \in \mathbb{N} : \|P_N U P_{i-1}^{\perp}\| \leq 1/(32K\sqrt{s})\}.$$

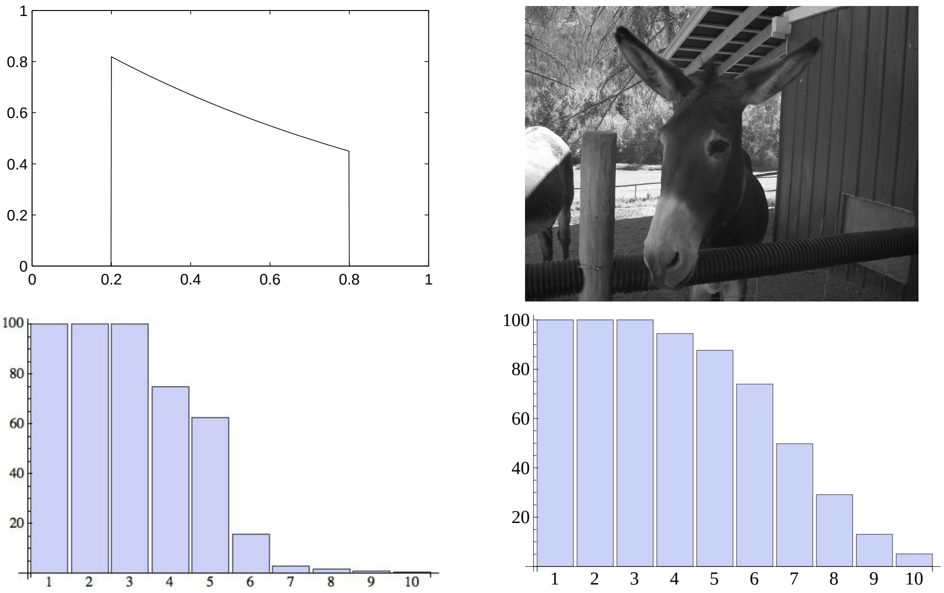


Figure 3: Top row: test functions. Bottom row: percentage of wavelet coefficients at each scale which are greater than 10^{-3} in magnitude.

Suppose that $\xi \in l^1(\mathbb{N})$ *is a minimizer of (4.5). Then, with probability exceeding* $1-\epsilon$,

$$\|\xi - x\| \leq C \cdot \left(1+\sqrt{K}\right) \cdot \left(\delta \cdot \left(1 + L \cdot \sqrt{s}\right) + \sigma_{\mathbf{s},\mathbf{M}}(f)\right), \tag{4.11}$$

for some constant C*, where* $\sigma_{\mathbf{s},\mathbf{M}}(f)$ *is as in (3.2), and* $L = \sqrt{6K}\left(\frac{3}{2} + \sqrt{1 + \frac{\log_2(\epsilon^{-1})+1}{\log_2(4K\tilde{M}\sqrt{s})}}\right)$. *If* $m_k = N_k - N_{k-1}$ *for* $1 \leq k \leq r$ *then this holds with probability* 1.

This theorem is very similar to that of the finite-dimensional case (Theorem 4.7), except for two key differences. First, as in the two-level setting, it requires the balancing property (see Section 4.1.2). Second, the final logarithm involves the quantity $\tilde{M}$. Note that $\tilde{M}$ is finite, since U is an isometry. In the case of Fourier sampling with wavelets, for example, $\tilde{M} = \mathcal{O}(KN)$.

5 Main consequences

We now in a position to discuss the main consequences of our theorems. This is the content of Sections 5.2–5.4. Before doing so, however, we first need to explain the relevance of our framework.

5.1 Real-life images are asymptotically sparse in wavelets

The reader may at this stage be asking whether or not the new definitions of asymptotic sparsity and compressibility are reasonable in practice. The answer to this is emphatically yes: natural, real-life images possess exactly these types of sparsity patterns in their wavelet coefficients.

In Figure 3 we tabulate the number of significant wavelet coefficients at each wavelet scale for two test functions. At coarse scales, there is very little sparsity. However, sparsity rapidly increases as the wavelet scale gets finer, demonstrating asymptotic sparsity in both cases.

This is not a new insight. Take, for example, a piecewise constant function f of one variable with a jump at an irrational point. It is well known that the number of nonzero wavelet coefficients at level k scales like the logarithm of the total number of coefficients at that level. More generally, the wavelet coefficients of natural images possess a sparse tree structure [16]. This fits exactly into the notion of asymptotically sparse in levels.

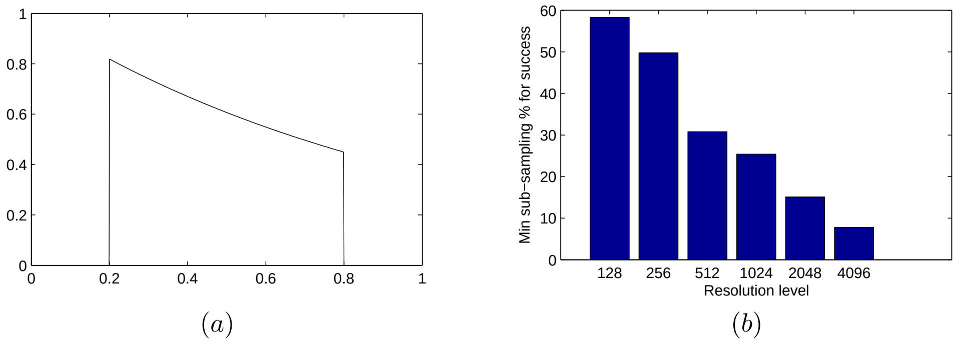


Figure 4: (a): the 1D phantom (5.1). (b): the minimum subsampling percentage p for success of the two-level sampling scheme.

5.2 First consequence: the success of compressed sensing is resolution dependent

We now discuss the first, and main, consequence of our theorems: *the success of compressed sensing is resolution dependent.*

Let us explain this phenomenon in more detail. Suppose the resolution $M := M_r$ in the wavelet domain is relatively low. The image f, although asymptotically sparse, is not particularly sparse at this resolution. Therefore regardless of how we choose to recover f, we will require a relatively large proportion of measurements. Taking too few measurements will give to a poor quality reconstruction, and as discussed in Section 1, lead to erroneous conclusions about the usefulness of compressed sensing.

On the other hand, when M is large, asymptotic sparsity kicks in and f becomes increasingly sparse at finer wavelet scales. As shown by our theorems, multilevel sampling allows one to exploit this property and recover f using a much smaller percentage of total samples than in the low resolution case. Hence compressed sensing becomes substantially more effective at higher resolutions.

Example 5.1 Consider the setup of Example 2.2 with the 1D phantom (visualized in Figure 4(a)) given by

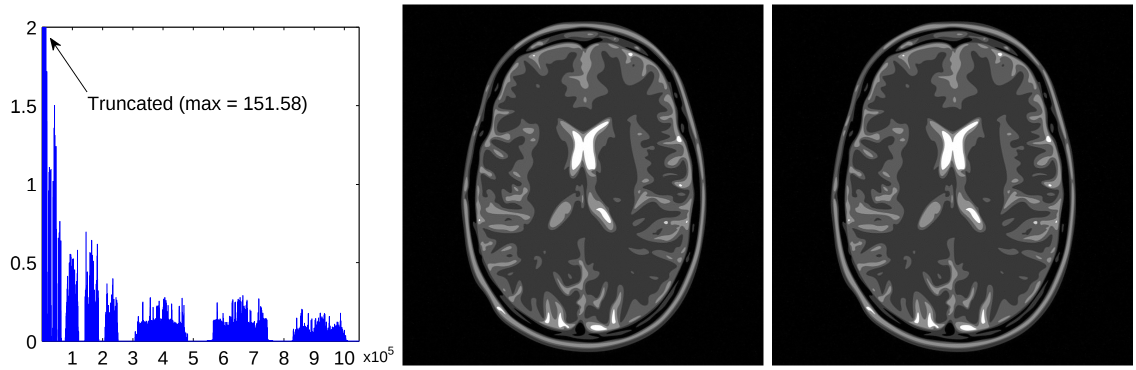
$$f(t) = \mathrm{e}^{-t}\chi_{[0.2,0.8]}(t), \quad t \in [0,1]. \tag{5.1}$$

We sample this function by measuring its Fourier coefficients, and take the sparsity system $\{\varphi_j\}_{j\in\mathbb{N}}$ to be the orthonormal basis of Haar wavelets on $[0,1]$. We then take a two-level sampling scheme with $p/2\%$ fixed samples and $p/2\%$ random samples, where p is the total subsampling percentage. In our experiment, we search for the smallest value of p such that the two-level sampling scheme succeeds: namely, it gives an error that is no worse than that of full sampling (by full sampling, we mean the scheme that uses all possible samples in the resolution range, i.e. that based on $\Omega = \{1,\ldots,N\}$).

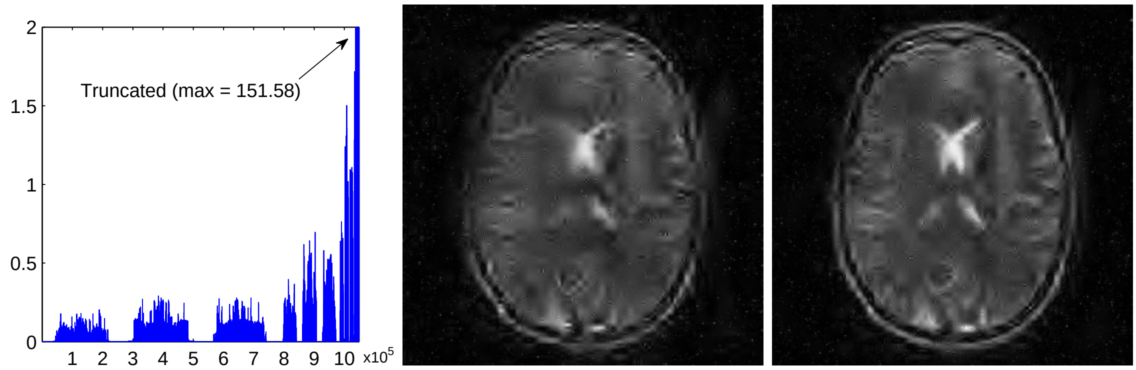
In Figure 4(b) we plot p against the resolution level N. For each N we have the same sampling device (Fourier samples), the same recovery algorithm (compressed sensing with two-level sampling) and the same function to recover. However, the difference between low resolution ($N = 128$) and high resolution ($N = 4096$) is dramatic. For the former, we require nearly 60% of the samples, whereas this value drops to less than 8% as we move to the higher resolution.

5.3 Second consequence: the optimal sampling strategy depends on the signal structure

Theorems 4.7 and 4.8 demonstrate that the required sampling density at the k^{th} level is determined by the quantities $\mu_{\mathbf{N},\mathbf{M}}(k,l)$, μ_{N_l}, s_k and S_k. Hence the optimal sampling strategy is completely dependent on the *signal structure*. Specifically, it is determined by the widths $M_k - M_{k-1}$ of the sparsity bands, and the sparsity levels s_k within them.



(a) Wavelet coefficients and subsampling reconstructions from 10% of Fourier coefficients.



(b) Reversed wavelet coefficients and the reconstructions from the reversed coefficients.

Figure 5: *Left*: Daubechies 4 wavelet coefficients of the analytic phantom (see Section 7.1). *Center*: reconstruction from 10% Fourier coefficients from a multi-level subsampling scheme (see Section 7.2). *Right*: reconstruction from 10% Fourier coefficients from a subsampling scheme using a continous power law (see Section 7.3). Note that had the RIP held, the upper and lower figures would have been identical.

This phenomenon can be illustrated by the following example:

Example 5.2 In Figure 5 we consider the reconstruction of a phantom brain image (see Section 7.1 for a discussion). As can be seen in Figure 5(a) this image is asymptotically sparse in wavelets, and can therefore be recovered well using either a multilevel or power law subsampling scheme. Now suppose we perform the following experiment. We reverse the ordering of the first 1024 wavelet coefficients (Figure 5(b)) in order to obtain a new function $\tilde{f}$, apply the same sampling patterns to recover $\tilde{f}$ from its Fourier measurements. Having done this we once more reverse the order of the (reconstructed) wavelet coefficients so as to obtain a reconstruction of f. The result of this process is shown in Figure 5(b). As is evident, this gives as completely meaningless reconstruction. In particular, the same sampling pattern, the same total sparsity, but different a signal structure yields highly contrasting results.

5.4 Third consequence: no Restricted Isometry Propery (RIP)

The informed reader will have noticed that our main theorems do not require involve the Restricted Isometry Property (RIP); a standard technique in compressed sensing theory [25, 26, 27]. Instead, our results are based solely on the coherence (or asymptotic/local coherence) of the matrix U. In this sense, our work continues the development of the *RIPless* theory of compressed sensing, as introduced by Candès & Plan in [12].

However, since the RIP-based techniques are well established, it is worth posing the following question: is the RIP relevant for imaging problems? That is to say, for realistic parameter

values (i.e. problem resolution, subsampling percentage, and sparsity) used in applications can one reasonably expect the RIP to hold? It is our belief that the answer to this is no. Had such matrices satisfied the RIP, one could permute the wavelet coefficients of an image and still recover it with the same accuracy. Figure 5 shows that this is not the case: it is clear that there is no RIP. The experiment in Figure 5 was repeated up to large subsampling percentage (50%) with many different images, but with the same outcome.

In view of this, the third conclusion of our work is that the RIP is of limited value in analysing compressive imaging strategies. Coherence, on the other hand, is both a relevant and powerful tool for understanding recoverability in this setting.

6 Proofs

The proofs rely on some key propositions from which one can deduce the main theorems. The main work is to prove these proposition, and that will be done subsequently.

6.1 Key results

Proposition 6.1. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ and suppose that Δ and $\Omega = \Omega_1 \cup \ldots \cup \Omega_r$ (where the union is disjoint) are subsets of $\mathbb{N}$. Let $x_0 \in \mathcal{H}$ and $z \in \mathrm{ran}(P_\Omega U)$ be such that $\|z\| \leq \delta$ for $\delta \geq 0$. Let $M \in \mathbb{N}$ and $y = P_\Omega U x_0 + z$ and $y_M = P_\Omega U P_M x_0 + z$. Suppose that $\xi \in \mathcal{H}$ and $\xi_M \in \mathcal{H}$ satisfiy*

$$\|\xi\|_{l^1} = \inf_{\eta \in \mathcal{H}} \{\|\eta\|_{l^1} : \|P_\Omega U \eta - y\| \leq \delta\}. \tag{6.1}$$

$$\|\xi_M\|_{l^1} = \inf_{\eta \in \mathcal{H}} \{\|\eta\|_{l^1} : \|P_\Omega U \eta - y_M\| \leq \delta\}. \tag{6.2}$$

If there exists a vector $\rho = U^* P_\Omega w$ such that

(i) $\|(P_\Delta U^* \left(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}\right) U P_\Delta|_{P_\Delta(\mathcal{H})})^{-1}\| \leq \frac{4}{3}$

(i)* $\|P_\Delta U^* \left(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}\right)\| \leq \sqrt{\frac{5}{4q}}$

(ii) $\|P_\Delta \rho - \mathrm{sgn}(P_\Delta x_0)\| \leq \frac{q}{8}$.

(iii) $\|P_\Delta^\perp \rho\|_{l^\infty} \leq \frac{1}{2}$

(iv) $\|w\| \leq L \cdot \sqrt{|\Delta|}$

then we have that

$$\|\xi - x_0\| \leq C \cdot \left(1 + \frac{1}{\sqrt{q}}\right) \cdot \left(\delta \cdot \left(1 + L \cdot \sqrt{s}\right) + \|P_\Delta^\perp x_0\|_{l^1}\right),$$

for some constant C. Also, if (iii) is replaced by $\|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq \frac{1}{2}$ then

$$\|\xi_M - x_0\| \leq C \cdot \left(1 + \frac{1}{\sqrt{q}}\right) \cdot \left(\delta \cdot \left(1 + L \cdot \sqrt{s}\right) + \|P_M P_\Delta^\perp x_0\|_{l^1}\right). \tag{6.3}$$

Proof. Suppose that there exists a vector ρ, constructed with $y_0 = P_\Delta x_0$, satisfying (i), (i*)-(iv). Let ξ be a solution to (6.1) and let $h = \xi - x_0$. Let $A_\Delta = P_\Delta U^* \left(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}\right) U P_\Delta|_{P_\Delta(\mathcal{H})}$. Then, the following holds:

$$\begin{aligned}
\|P_\Delta h\| &= \|A_\Delta^{-1} A_\Delta P_\Delta h\| \\
&\leq \|A_\Delta^{-1}\| \|P_\Delta U^* \left(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}\right) U (I - P_\Delta^\perp) h\| \\
&\leq \frac{4}{3} \|P_\Delta U^* \left(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}\right)\| \left(\|P_\Omega U h\| + \|P_\Delta^\perp h\|\right) \\
&\leq \frac{2\sqrt{5}}{3\sqrt{q}} \left(2\delta + \|P_\Delta^\perp h\|_{l^1}\right),
\end{aligned} \tag{6.4}$$

where the second inequality follows from (i) and the last inequality follows from (i*). We will now obtain a bound for $\|P_\Delta^\perp h\|_{l^1}$. In particular,

$$\begin{aligned}\|h+x_0\|_{l^1} &= \|P_\Delta h + P_\Delta x_0\|_{l^1} + \|P_\Delta^\perp (h+x_0)\|_{l^1} \\ &\geq \operatorname{Re}\langle P_\Delta h, \operatorname{sgn}(P_\Delta x_0)\rangle + \|P_\Delta x_0\|_{l^1} + \|P_\Delta^\perp h\|_{l^1} - \|P_\Delta^\perp x_0\|_{l^1} \\ &\geq \operatorname{Re}\langle P_\Delta h, \operatorname{sgn}(P_\Delta x_0)\rangle + \|x_0\|_{l^1} + \|P_\Delta^\perp h\|_{l^1} - 2\|P_\Delta^\perp x_0\|_{l^1}\end{aligned} \tag{6.5}$$

Since $\|x_0\|_{l^1} \geq \|h+x_0\|_{l^1}$, we have that

$$\|P_\Delta^\perp h\|_{l^1} \leq |\langle P_\Delta h, \operatorname{sgn}(P_\Delta x_0)\rangle| + 2\|P_\Delta^\perp x_0\|_{l^1}. \tag{6.6}$$

We will use this equation later on in the proof, but before we do that observe that some basic adding and subtracting yields

$$\begin{aligned}|\langle P_\Delta h, \operatorname{sgn}(x_0)\rangle| &\leq |\langle P_\Delta h, \operatorname{sgn}(P_\Delta x_0) - P_\Delta \rho\rangle| + |\langle h, \rho\rangle| + \left|\langle P_\Delta^\perp h, P_\Delta^\perp \rho\rangle\right| \\ &\leq \|P_\Delta h\| \|\operatorname{sgn}(P_\Delta x_0) - P_\Delta \rho\| + |\langle P_\Omega U h, w\rangle| + \|P_\Delta^\perp h\|_{l^1}\|P_\Delta^\perp \rho\|_{l^\infty} \\ &\leq \frac{q}{8}\|P_\Delta h\| + 2L\delta\sqrt{s} + \frac{1}{2}\|P_\Delta^\perp h\|_{l^1} \\ &\leq \frac{\sqrt{5q}}{12}\left(2\delta + \|P_\Delta^\perp h\|_{l^1}\right) + 2L\delta\sqrt{s} + \frac{1}{2}\|P_\Delta^\perp h\|_{l^1}\end{aligned} \tag{6.7}$$

where the last inequality utilises (6.4) and the penultimate inequality follows from properties (ii), (iii) and (iv) of the dual vector ρ. Combining with (6.6) gives that

$$\|P_\Delta^\perp h\|_{l^1} \leq \delta\left(\frac{2\sqrt{5q}}{3} + 8L\sqrt{s}\right) + 8\|P_\Delta^\perp x_0\|_{l^1}. \tag{6.8}$$

Thus, (6.4) and (6.8) yields:

$$\begin{aligned}\|h\| &\leq \frac{4}{3\sqrt{q}}\left(2\delta + \|P_\Delta^\perp h\|_{l^1}\right) + \|P_\Delta^\perp h\|_{l^1} \\ &\leq \frac{8\delta}{3\sqrt{q}} + \delta\left(\frac{4}{3\sqrt{q}} + 1\right)\left(\frac{2\sqrt{5q}}{3} + 8L\sqrt{s}\right) + \left(\frac{32}{3\sqrt{q}} + 8\right)\|P_\Delta^\perp x_0\|_{l^1}\end{aligned} \tag{6.9}$$

□

The next two propositions give sufficient conditions for Proposition 6.1 to be true. But before we state them we need to define the following.

Definition 6.2. *Let U be an isometry of either $\mathbb{C}^{N\times N}$ or $\mathcal{B}(l^2(\mathbb{N}))$. For $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$, $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \ldots < N_r$ and $1 \leq M_1 < \ldots < M_r$, $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$ and $1 \leq k \leq r$, let*

$$\kappa_{\mathbf{N},\mathbf{M}}(k,l) = \max_{\eta \in \Theta} \|P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}} \eta\|_{l^\infty} \cdot \sqrt{\mu(P_{N_{k-1}}^\perp U)}.$$

where

$$\Theta = \{\eta : \|\eta\|_{l^\infty} \leq 1, |\operatorname{supp}(P_{M_l}^{M_{l-1}}\eta)| = s_l,\, l = 1, \ldots, r\},$$

and $N_0 = M_0 = 0$.

Proposition 6.3. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry and $x \in l^1(\mathbb{N})$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme, where $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$. Let $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$, $M_1 < \ldots < M_r$, and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, be any pair such that the following holds:*

(i) The parameters

$$N := N_r, \quad K := \max_{k=1,\ldots,r}\left\{\frac{N_k - N_{k-1}}{m_k}\right\},$$

satisfy the weak balancing property with respect to U, $M := M_r$ and $s := s_1 + \ldots + s_r$;

(ii) *for $\epsilon > 0$ and $1 \leq k \leq r$,*

$$1 \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \frac{N_k - N_{k-1}}{m_k} \cdot (\nu_{k,1} + \ldots + \nu_{k,r}) \cdot \log\left(KM\sqrt{s}\right), \tag{6.10}$$

where

$$\nu_{k,l} = \min(\mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot s_l, \kappa_{\mathbf{N},\mathbf{M}}(k,l));$$

(iii)

$$m_k \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \hat{m}_k \cdot \log\left(KM\sqrt{s}\right), \tag{6.11}$$

where $\hat{m}_k$ satisfies

$$1 \geq \left(\frac{N_1 - N_0}{\hat{m}_1} - 1\right) \cdot \mu_{N_0} \cdot \tilde{s}_1 + \ldots + \left(\frac{N_r - N_{r-1}}{\hat{m}_r} - 1\right) \cdot \mu_{N_{r-1}} \cdot \tilde{s}_r,$$

where

$$\tilde{s}_1 + \ldots + \tilde{s}_r \leq s_1 + \ldots + s_r, \qquad \tilde{s}_k \leq S_k(s_1, \ldots, s_r),$$

and S_k is defined in (4.6).

Then (i), (i)-(iv) in Proposition 6.1 follow with probability exceeding $1-\epsilon$, with (iii) is replaced by $\|P_M P_\Delta^\perp \rho\|_{l^\infty} \leq \frac{1}{2}$ and L in (iv) is given by*

$$L = \sqrt{6K}\left(\frac{3}{2} + \sqrt{1 + \frac{\log_2(6\epsilon^{-1}) + 1}{\log_2(4KM\sqrt{s})}}\right). \tag{6.12}$$

If $m_k = N_k - N_{k-1}$ for all $1 \leq k \leq r$ then (i), (i)-(iv) follow with probability one.*

Proposition 6.4. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry and $x \in l^1(\mathbb{N})$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme, where $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$. Let $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$, $M_1 < \ldots < M_r$, and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, be any pair such that the following holds:*

(i) *The parameters N and K (as in Proposition 6.3) satisfy the strong balancing property with respect to U, $M = M_r$ and $s := s_1 + \ldots + s_r$;*

(ii) *for $\epsilon > 0$ and $1 \leq k \leq r$,*

$$1 \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \frac{N_k - N_{k-1}}{m_k} \cdot (\nu_{k,1} + \ldots + \nu_{k,r}) \cdot \log\left(K\tilde{M}\sqrt{s}\right), \tag{6.13}$$

where $\nu_{k,l}$ is as in Proposition 6.3;

(iii)

$$m_k \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \hat{m}_k \cdot \log\left(K\tilde{M}\sqrt{s}\right), \tag{6.14}$$

where $\tilde{M} = \min\{i \in \mathbb{N} : \|P_N U P_{i-1}^\perp\| \leq 1/(K32\sqrt{s})\}$

Then (i), (i)-(iv) in Proposition 6.1 follow with probability exceeding $1-\epsilon$ with L as in (6.12). If $m_k = N_k - N_{k-1}$ for all $1 \leq k \leq r$ then (i), (i*)-(iv) follow with probability one.*

Lemma 6.5 (Bounds for $\kappa_{\mathbf{N},\mathbf{M}}(k,l)$ and S_k)**.** *For $k = 1, \ldots, r-1$, if*

$$\|P_{N_k}^\perp U P_{M_k}\| \leq \frac{\gamma}{\sqrt{s_1 + \cdots + s_k}}, \quad \gamma \leq \frac{2}{5}$$

then

$$\kappa_{\mathbf{N},\mathbf{M}}(k,l) \leq \gamma \cdot \sqrt{\mu(P_{N_{k-1}}^\perp U)}, \quad l < k \tag{6.15}$$

and

$$S_{k+1} \leq 2\left(s_{k+1} + \ldots + s_r\right)$$

Proof. Let $\|P_{N_k}^{\perp}UP_{M_k}\| = \gamma_k$ and note that

$$\kappa_{\mathbf{N},\mathbf{M}}(k,l) \leq \gamma_l \cdot \sqrt{s_1 + \cdots + s_l} \cdot \sqrt{\mu(P_{N_{k-1}}^{\perp}U)}, \quad l < k$$

and

$$\begin{aligned} S_{k+1} &= \max_{\eta \in \Theta} \|P_{N_{k+1}}^{N_k} U\eta\|^2 \\ &\leq \max_{\eta \in \Theta} \left(\|P_{N_k}^{\perp}UP_{M_k}\eta\| + \|P_{N_{k+1}}UP_{M_k}^{\perp}\eta\|\right)^2 \\ &\leq \left(\gamma_k\sqrt{s_1 + \ldots + s_k} + \sqrt{s_{k+1} + \ldots + s_r}\right)^2 \end{aligned}$$

□

We are now ready to prove the main theorems.

Proof of Theorems 4.2 and 4.4 . It is straightforward to show that Theorem 4.2 follows from Theorem 4.4, thus we will concentrate on proving the latter. Note that Proposition 6.3 applied to a two-level sampling scheme $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$, where $\mathbf{N} = (N_1, N_2)$ and $\mathbf{m} = (m_1, m_2)$ with $m_1 = M_1$ and $m_2 = m$, and $x \in l^1(\mathbb{N})$ which is $(\mathbf{s},\mathbf{M})$-sparse, where $r = 2$, $\mathbf{s} = (M_1, s_2)$, $\mathbf{M} = (M_1, M_2)$, and $N_1, N_2, m_1, m_2 \in \mathbb{N}$ are such that

$$N = N_2, \quad K = \max\left\{\frac{N_2 - N_1}{m_2}, \frac{N_1}{m_1}\right\}$$

satisfy the weak balancing property with respect to U, $M = M_2$ and $s = M_1 + s_2$, yield the following: (i), (i*)-(iv) in Proposition 6.1 follow with probability exceeding $1-\epsilon$, with (iii) replaced by $\|P_M P_{\Delta}^{\perp}\rho\|_{l^\infty} \leq \frac{1}{2}$ and L in (iv) is given by (6.12), if

$$1 \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \frac{N - N_1}{m_2} \cdot (\nu_{2,1} + \nu_{2,2}) \cdot \log\left(KM\sqrt{s}\right), \tag{6.16}$$

$$m_2 \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \hat{m}_2 \cdot \log\left(KM\sqrt{s}\right), \tag{6.17}$$

where $\hat{m}_k$ satisfies

$$1 \geq \left(\frac{N_r - N_{r-1}}{\hat{m}_r} - 1\right) \cdot \mu_{N_1} \cdot \tilde{s}_2,$$

where $\tilde{s}_2 \leq S_2(s_1, s_2)$, and S_k is defined in (4.6). But via the assumption that $\|P_{N_1}^{\perp}UP_{M_1}\| \leq \gamma/\sqrt{M_1}$ for some $\gamma \leq 2/5$ and Lemma 6.5 we observe that (4.4) implies (6.16) and (6.17). Thus, the theorem now follows from Proposition 6.1. □

The proof of Theorem 4.5 is almost verbatim the proof of Theorem 4.4 with the exception that all references to Proposition 6.3 is replaced by Proposition 6.4. We omit the details.

Proof of Theorem 4.7 and Theorem 4.8. It is straightforward that Theorem 4.7 follows from Theorem 4.8 and a direct application of Proposition 6.4 and Proposition 6.1 completes the proof. □

What is left is now to prove Propositions 6.3 and 6.4, and that will be done in the next sections.

6.2 Preliminaries

Before we start on the rather long way to prover the propositions, let us recall one of the monumental theorems in probability theory that will come in handy later on.

Theorem 6.6. *(Talagrand [43]) There exists a number K with the following property. Consider n independent random variables X_i valued in a measurable space Ω. Consider a (countable) class $\mathcal{F}$ of measurable functions on Ω. Consider the random variable $Z = \sup_{f\in\mathcal{F}} \sum_{i\leq n} f(X_i)$. Consider*

$$S = \sup_{f\in\mathcal{F}} \|f\|_\infty, \qquad V = \sup_{f\in\mathcal{F}} \mathbb{E}\left(\sum_{i\leq n} f(X_i)^2\right).$$

If $\mathbb{E}(f(X_i)) = 0$ for all $f \in \mathcal{F}$ and $i \leq n$, then, for each $t > 0$, we have

$$\mathbb{P}(|Z - \mathbb{E}(Z)| \geq t) \leq 3 \exp\left(-\frac{1}{K}\frac{t}{S}\log\left(1 + \frac{tS}{V + S\mathbb{E}(\overline{Z})}\right)\right),$$

where $\overline{Z} = \sup_{f\in\mathcal{F}} |\sum_{i\leq n} f(X_i)|$.

Equipped with this sledge hammer we can now get to work. We will present the following theorem as well as two technical propositions that will serve as the main tools in the proofs of the main propositions. A crucial tool in the proofs of the propositions is the Bernoulli sampling model. We will use the notation $\{a,\dots,b\} \supset \Omega \sim \mathrm{Ber}(q)$, where $a < b$ $a, b \in \mathbb{N}$, when $\Omega = \{k : \delta_k = 1\}$ and $\{\delta_k\}_{k=1}^N$ is a sequence of Bernoulli variables with $\mathbb{P}(\delta_k) = 1 = q$

Definition 6.7. *Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1,\dots,N_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \ldots < N_r$, $\mathbf{m} = (m_1,\dots,m_r) \in \mathbb{N}^r$, with $m_k \leq N_k - N_{k-1}$, $k = 1,\dots,r$, and suppose that*

$$\Omega_k \subseteq \{N_{k-1}+1,\dots,N_k\}, \quad \Omega_k \sim \mathrm{Ber}\left(\frac{m_k}{N_k - N_{k-1}}\right), \quad k = 1,\dots,r,$$

where $N_0 = 0$. We refer to the set

$$\Omega = \Omega_{\mathbf{N},\mathbf{m}} := \Omega_1 \cup \ldots \cup \Omega_r.$$

as an $(\mathbf{N},\mathbf{m})$-multilevel Bernoulli sampling scheme.

Theorem 6.8. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel Bernoulli sampling scheme, where $\mathbf{N} = (N_1,\dots,N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1,\dots,m_r) \in \mathbb{N}^r$. Consider $(\mathbf{s},\mathbf{M})$, where $\mathbf{M} = (M_1,\dots,M_r) \in \mathbb{N}^r$, $M_1 < \ldots < M_r$, and $\mathbf{s} = (s_1,\dots,s_r) \in \mathbb{N}^r$, and let*

$$\Delta = \Delta_1 \cup \ldots \cup \Delta_r, \qquad \Delta_k \subset \{M_{k-1},\dots,M_k\}, \qquad |\Delta_k| = s_k$$

where $M_0 = 0$. If $\|P_{M_r}U^*P_{N_r}UP_{M_r} - P_{M_r}\| \leq 1/8$ then, for $\gamma \in (0,1)$,

$$\mathbb{P}(\|P_\Delta U^*(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r})UP_\Delta - P_\Delta\| \geq 1/4) \leq \gamma, \tag{6.18}$$

where $q_k = m_k/(N_k - N_{k-1})$, provided that

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot (\nu_{k,1} + \ldots + \nu_{k,r}) \cdot \left(\log\left(\gamma^{-1}s\right) + 1\right), \quad \nu_{k,l} = \min(\mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot s_l, \kappa_{\mathbf{N},\mathbf{M}}(k,l)). \tag{6.19}$$

In addition, if $q = 1$, then

$$\mathbb{P}(\|P_\Delta U^*(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r})UP_\Delta - P_\Delta\| \geq 1/4) = 0.$$

To prove this theorem we deliberately avoid the use of the Matrix Bernstein inequality [28]. The reason is that this inequality is dimension dependent and hence yields dimension dependent estimates if used to prove the main theorems (see the dimension dependent bounds on the probability in [12] for example). As we are proving theorems for infinite-dimensional problems it is most unnatural to use dimension dependent result, and we therefore rely on Talagrand's dimension free theorem. However, before we can prove our theorem, we need a technical lemma.

Lemma 6.9. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ with $\|U\| \leq 1$, and consider the setup in Theorem 6.18. Let $\{\delta_j\}_{j=1}^N$ be random independent Bernoulli variables with $\mathbb{P}(\delta_j = 1) = q_j$, $q_j = m_k/(N_k - N_{k-1})$ and $j \in \{N_{k-1}+1,\dots,N_k\}$, and define $Z = \sum_{j=1}^N Z_j$, $Z_j = \left(q_j^{-1}\delta_j - 1\right)\eta_j \otimes \bar{\eta}_j$ and $\eta_j = P_\Delta U^* e_j$. Then*

$$\mathbb{E}\left(\|Z\|\right)^2 \leq 24\log(|\Delta|) \max_{1\leq j\leq N}\left\{q_j^{-1}\|\eta_j\|^2\right\},$$

when

$$\log(|\Delta|)^{-1} \geq 9 \max_{1\leq j\leq N}\left\{q_j^{-1}\|\eta_j\|^2\right\}.$$

The proof of this lemma is essentially reworking a result by Rudelson [40], but we need to include it here as the setup deviates from similar results in [13] and [2].

Lemma 6.10. *(Rudelson) Let $\eta_1,\dots,\eta_M \in \mathbb{C}^n$ and let $\varepsilon_1,\dots \varepsilon_M$ be independent Bernoulli variables taking values $1,-1$ with probability $1/2$. Then*

$$\mathbb{E}\left(\left\|\sum_{i=1}^{M} \varepsilon_i \bar{\eta}_i \otimes \eta_i\right\|\right) \leq \frac{3}{2}\sqrt{\log(n)}\max_{i\leq M}\|\eta_i\|\sqrt{\left\|\sum_{i=1}^{M}\bar{\eta}_i\otimes\eta_i\right\|}$$

Lemma 6.10 is often referred to as Rudelson's Lemma [40], however, we are relying on a complex version that was proven by Tropp [44], where also the constant 3/2 was established.

Proof of Lemma 6.9. We we start by observing that by letting $\tilde{\delta} = \{\tilde{\delta}_j\}_{j=1}^N$ be independent copies of $\delta = \{\delta_j\}_{j=1}^N$. Then, since $\mathbb{E}(Z) = 0$,

$$\begin{aligned}
\mathbb{E}_\delta\left(\|Z\|\right) &= \mathbb{E}_\delta\left(\left\|Z - \mathbb{E}_{\tilde{\delta}}\left(\sum_{j=1}^{N}\left(q_j^{-1}\tilde{\delta}_j - 1\right)\eta_j\otimes\bar{\eta}_j\right)\right\|\right) \\
&\leq \mathbb{E}_\delta\left(\mathbb{E}_{\tilde{\delta}}\left(\left\|Z - \sum_{j=1}^{N}\left(q_j^{-1}\tilde{\delta}_j - 1\right)\eta_j\otimes\bar{\eta}_j\right\|\right)\right),
\end{aligned} \tag{6.20}$$

by Jensen's inequality. Let $\varepsilon = \{\varepsilon_j\}_{j=1}^N$ be a sequence of Bernoulli variables taking values ± 1 with probability 1/2. Then, by (6.20), symmetry, Fubini's Theorem and the triangle inequality, it follows that

$$\begin{aligned}
\mathbb{E}_\delta\left(\|Z\|\right) &\leq \mathbb{E}_\varepsilon\left(\mathbb{E}_\delta\left(\mathbb{E}_{\tilde{\delta}}\left(\left\|\sum_{j=1}^{N}\varepsilon_j\left(q_j^{-1}\delta_j - q_j^{-1}\tilde{\delta}_j\right)\eta_j\otimes\bar{\eta}_j\right\|\right)\right)\right) \\
&\leq 2\mathbb{E}_\delta\left(\mathbb{E}_\varepsilon\left(\left\|\sum_{j=1}^{N}\varepsilon_j q_j^{-1}\delta_j\eta_j\otimes\bar{\eta}_j\right\|\right)\right).
\end{aligned} \tag{6.21}$$

Note that the setup in (6.21) is now ready for the use of Rudelson's Lemma (Lemma 6.10). However, as specified before, it is the complex version that is crucial here. Now, by Lemma 6.10 we get that

$$\mathbb{E}_\varepsilon\left(\left\|\sum_{j=1}^{N}\varepsilon_j q_j^{-1}\delta_j\eta_j\otimes\bar{\eta}_j\right\|\right) \leq \frac{3}{2}\sqrt{\log(s)}\max_{1\leq j\leq N} q_j^{-1/2}\|\eta_j\|\sqrt{\left\|\sum_{j=1}^{N}q_j^{-1}\delta_j\eta_j\otimes\bar{\eta}_j\right\|}. \tag{6.22}$$

And hence, by using (6.21) and (6.22), it follows that

$$\mathbb{E}_\delta\left(\|Z\|\right) \leq 3\sqrt{\log(s)}\max_{1\leq j\leq N} q_j^{-1/2}\|\eta_j\|\sqrt{\mathbb{E}_\delta\left(\left\|Z + \sum_{j=1}^{N}\eta_j\otimes\bar{\eta}_j\right\|\right)}.$$

Thus, by using the easy calculus fact that if $r > 0$, $c \leq 1$ and $r \leq c\sqrt{r+1}$ then we have that $r \leq c(1+\sqrt{5})/2$, and the fact that U is an isometry (so that $\|\sum_{j=1}^N \eta_j\otimes\bar{\eta}_j\| \leq 1$), it is easy to see that the claim follows. □

We are now ready to prove Theorem 6.8

Proof of Theorem 6.8. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables as defined in Lemma 6.9 and define $Z = \sum_{j=1}^N Z_j$, $Z_j = \left(q_j^{-1}\delta_j - 1\right)\eta_j\otimes\bar{\eta}_j$ with $\eta_j = P_\Delta U^* e_j$. Now observe that

$$P_\Delta U^*(q_1^{-1}P_{\Omega_1}\oplus\dots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta = \sum_{j=1}^{N} q_j^{-1}\delta_j\eta_j\otimes\bar{\eta}_j, \quad P_\Delta U^* P_N U P_\Delta = \sum_{j=1}^{N}\eta_j\otimes\bar{\eta}_j. \tag{6.23}$$

Thus, it follows that

$$\begin{aligned}\|P_\Delta U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_1^{-1}P_{\Omega_1})UP_\Delta - P_\Delta\| &\leq \|Z\| + \|(P_\Delta U^* P_N U P_\Delta - P_\Delta)\| \\ &\leq \|Z\| + \frac{1}{8},\end{aligned} \tag{6.24}$$

by the assumed weak balancing property. Thus, to prove the assertion we need to estimate $\|Z\|$, and Talagrand's Theorem 6.6 will be our main tool. Note that clearly, since Z is self-adjoint, we have that $\|Z\| = \sup_{\zeta\in\mathcal{G}} |\langle Z\zeta,\zeta\rangle|$, where $\mathcal{G}$ is a countable set of vectors in the unit ball of $P_\Delta(\mathcal{H})$. Define, for $\zeta \in \mathcal{G}$, the mappings

$$\hat{\zeta}_1(T) = \langle T\zeta,\zeta\rangle, \quad \hat{\zeta}_2(T) = -\langle T\zeta,\zeta\rangle, \qquad T\in\mathcal{B}(\mathcal{H}).$$

In order to use Talagrand's Theorem 6.6 we restrict the domain $\mathcal{D}$ of the mappings ζ_i to

$$\mathcal{D} = \{T\in\mathcal{B}(\mathcal{H}) : \|T\| \leq \max_{1\leq j\leq N}\{q_j^{-1}\|\eta_j\|^2\}\}.$$

Let $\mathcal{F}$ denote the family of mappings $\hat{\zeta}_1,\hat{\zeta}_2$ for $\zeta\in\mathcal{G}$. Then $\|Z\| = \sup_{\hat{\zeta}_i\in\mathcal{F}}\hat{\zeta}(Z)$, and we have

$$|\hat{\zeta}_i(Z_j)| = \left|\left(q_j^{-1}\delta_j - 1\right)\right| |\langle(\eta_j\otimes\bar{\eta}_j)\,\zeta,\zeta\rangle| \leq \max_{1\leq j\leq N}\{q_j^{-1}\|\eta_j\|^2\}.$$

Thus, $Z_j\in\mathcal{D}$ for $1\leq j\leq N$ and $S := \sup_{\zeta\in\mathcal{F}}\|\hat{\zeta}\|_\infty = \max_{1\leq j\leq N}\{q_j^{-1}\|\eta_j\|^2\}$. Note that

$$q_j^{-1}\|\eta_j\|^2 = q_j^{-1}\langle P_\Delta U^* e_j, P_\Delta U^* e_j\rangle = q_j^{-1}\sum_{k=1}^r \langle P_{\Delta_k}U^*e_j, P_{\Delta_k}U^*e_j\rangle.$$

Also, note that an easy application of Holder's inequality gives the following (note that the l^1 and l^∞ bounds are finite because all the projections have finite rank),

$$\begin{aligned}&|\langle P_{\Delta_k}U^*e_j, P_{\Delta_k}U^*e_j\rangle| \leq \|P_{\Delta_k}U^*e_j\|_{l^1}\|P_{\Delta_k}U^*e_j\|_{l^\infty} \\ &\leq \|P_{\Delta_k}U^*P_{N_l}^{N_{l-1}}\|_{l^1\to l^1}\|P_{\Delta_k}U^*e_j\|_{l^\infty} \leq \|P_{N_l}^{N_{l-1}}UP_{\Delta_k}\|_{l^\infty\to l^\infty}\cdot\sqrt{\mu(P_{N_{l-1}}^{\perp}U)} = \kappa_{\mathbf{N},\mathbf{M}}(l,k),\end{aligned}$$

for $j\in\{N_{l-1}+1,\ldots,N_l\}$ and $l\in\{1,\ldots,r\}$. And finally observe that

$$|\langle P_{\Delta_k}U^*e_j, P_{\Delta_k}U^*e_j\rangle| \leq \mu(P_{N_l}^{N_{l-1}}UP_{M_k}^{M_{k-1}})\cdot s_k \leq \mu_{\mathbf{N},\mathbf{M}}(l,k)\cdot s_k.$$

Hence, it follows that

$$\|\eta_j\|^2 \leq \max_{1\leq k\leq r}(\nu_{k,1}+\ldots+\nu_{k,r}), \qquad \nu_{k,l} = \min(\mu_{\mathbf{N},\mathbf{M}}(k,l)\cdot s_l, \kappa_{\mathbf{N},\mathbf{M}}(k,l)) \tag{6.25}$$

and therefore $S\leq\max_{1\leq k\leq r} q_k^{-1}(\nu_{k,1}+\ldots+\nu_{k,r})$. Finally, note that by (6.25) and the reasoning above, it follows that

$$\begin{aligned}V &:= \sup_{\hat{\zeta}_i\in\mathcal{F}}\mathbb{E}\left(\sum_{j=1}^N\hat{\zeta}_i(Z_j)^2\right) \leq \sup_{\hat{\zeta}_i\in\mathcal{F}}\mathbb{E}\left(\sum_{j=1}^N\left(q_j^{-1}\delta_j-1\right)^2|\langle P_\Delta U^*e_j,\zeta\rangle|^4\right) \\ &\leq \max_{1\leq k\leq r}\|\eta_k\|^2\left(\frac{N_k-N_{k-1}}{m_k}-1\right)\sup_{\hat{\zeta}_i\in\mathcal{F}}\sum_{j=1}^N|\langle e_j, UP_\Delta\zeta\rangle|^2, \\ &\leq \max_{1\leq k\leq r}\frac{N_k-N_{k-1}}{m_k}(\nu_{k,1}+\ldots+\nu_{k,r})\|U\zeta\| = \max_{1\leq k\leq r}\frac{N_k-N_{k-1}}{m_k}(\nu_{k,1}+\ldots+\nu_{k,r}),\end{aligned} \tag{6.26}$$

where we used the fact that U is an Isometry to deduce that $\|U\|=1$. Also, by Lemma 6.9 and (6.25) , it follows that

$$\mathbb{E}\left(\|Z\|\right)^2 \leq 24\max_{1\leq k\leq r}\frac{N_k-N_{k-1}}{m_k}(\nu_{k,1}+\ldots+\nu_{k,r})\cdot\log(s) \tag{6.27}$$

when

$$1 \geq 9 \max_{1 \leq k \leq r} \frac{N_k - N_{k-1}}{m_k} (\nu_{k,1} + \ldots + \nu_{k,r}) \cdot \log(s). \tag{6.28}$$

Thus, by (6.24) and Talagrand's Theorem 6.6, it follows that

$$\begin{aligned}
&\mathbb{P}\left(\|P_\Delta U^*(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}) U P_\Delta - P_\Delta\| \geq 1/4\right) \\
&\leq \mathbb{P}\left(\|Z\| \geq \frac{1}{16} + \sqrt{24 \max_{1 \leq k \leq r} \frac{N_k - N_{k-1}}{m_k} (\nu_{k,1} + \ldots + \nu_{k,r}) \cdot \log(s)}\right) \\
&\leq K \exp\left(-\frac{1}{16K}\left(\max_{1 \leq k \leq r} \frac{N_k - N_{k-1}}{m_k} (\nu_{k,1} + \ldots + \nu_{k,r})\right)^{-1} \log\left(1 + 1/32\right)\right),
\end{aligned} \tag{6.29}$$

when m_k's are chosen such that the right hand side of (6.27) is less than or equal to $\frac{1}{16^2}$. This yields the first part of the theorem. The second claim of this theorem follows from the Balancing property. ☐

Proposition 6.11. *Let $U \in \mathcal{B}(l^2(\mathbb{N}))$ be an isometry. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel Bernoulli sampling scheme, where $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$. Consider $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$, $M_1 < \ldots < M_r$, and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, and let*

$$\Delta = \Delta_1 \cup \ldots \cup \Delta_r, \qquad \Delta_k \subset \{M_{k-1}, \ldots, M_k\}, \qquad |\Delta_k| = s_k$$

where $M_0 = 0$. Let $\beta \geq 1/4$.

(i) If

$$N := N_r, \quad K := \max_{k=1,\ldots,r} \left\{\frac{N_k - N_{k-1}}{m_k}\right\},$$

satisfy the weak balancing property with respect to U, $M := M_r$ and $s := s_1 + \ldots + s_r$, then, for $\xi \in \mathcal{H}$ and $t, \gamma > 0$, we have that

$$\mathbb{P}\left(\|P_M P_\Delta^\perp U^*(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}) U P_\Delta \xi\|_{l^\infty} > \beta \|\xi\|_{l^\infty}\right) \leq \gamma, \tag{6.30}$$

provided that

$$\frac{\beta}{\log\left(\frac{4}{\gamma}(M - s)\right)} \geq C\ \Lambda, \qquad \frac{\beta^2}{\log\left(\frac{4}{\gamma}(M - s)\right)} \geq C\ \Upsilon, \tag{6.31}$$

for some constant $C > 0$, where

$$\Lambda = \max_{\substack{1 \leq k \leq r \\ k \in \mathcal{G}}} \left\{\frac{N_k - N_{k-1}}{m_k} \cdot (\nu_{k,1} + \ldots + \nu_{k,r})\right\}, \quad \nu_{k,l} = \min(\mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot s_l, \kappa_{\mathbf{N},\mathbf{M}}(k,l)). \tag{6.32}$$

where $\mathcal{G} = \{k : m_k < N_k - N_{k-1}, k = 1, \ldots, r\}$ and

$$\Upsilon = \left(\frac{N_1 - N_0}{m_1} - 1\right) \cdot \mu_{N_0} \cdot \tilde{s}_1 + \ldots + \left(\frac{N_r - N_{r-1}}{m_r} - 1\right) \cdot \mu_{N_{r-1}} \cdot \tilde{s}_r, \tag{6.33}$$

for some $\{\tilde{s}_k\}_{k=1}^r$ such that

$$\tilde{s}_1 + \ldots + \tilde{s}_r \leq s_1 + \ldots + s_r, \qquad \tilde{s}_k \leq S_k(s_1, \ldots, s_r).$$

Moreover, if $q_k = 1$ for all $k = 1, \ldots, r$, then (6.31) is trivially satisfied for any $\gamma > 0$ and the left-hand side of (6.30) is equal to zero.

(ii) If N satisfies the strong Balancing Property with respect to U, M and s, then, for $\xi \in \mathcal{H}$ and $t, \gamma > 0$, we have that

$$\mathbb{P}\left(\|P_\Delta^\perp U^*(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}) U P_\Delta \xi\|_{l^\infty} > \beta \|\xi\|_{l^\infty}\right) \leq \gamma, \tag{6.34}$$

provided that

$$\frac{\beta}{\log\left(\frac{4}{\gamma}(\tilde{\theta}-s)\right)} \geq C\ \Lambda, \qquad \frac{\beta^2}{\log\left(\frac{4}{\gamma}(\tilde{\theta}-s)\right)} \geq C\ \Upsilon, \tag{6.35}$$

for some constant $C>0$, $\tilde{\theta}=\tilde{\theta}(\{q_k\}_{k=1}^r, 1/8, \{N_k\}_{k=1}^r, s, M)$ *and* Υ, Λ *as defined in (i) and*

$$\begin{aligned}&\tilde{\theta}(\{q_k\}_{k=1}^r, t, \{N_k\}_{k=1}^r, s, M) \\ &= \left|\left\{ i \in \mathbb{N} : \max_{\substack{\Gamma_1 \subset \{1,\dots,M\}, \quad |\Gamma_1|=s \\ \Gamma_{2,j}\subset\{N_{j-1}+1,\dots,N_j\}, \quad j=1,\dots,r}} \|P_{\Gamma_1}U^*(q_1^{-1}P_{\Gamma_{2,1}}\oplus\dots\oplus q_r^{-1}P_{\Gamma_{2,r}})Ue_i\| > \frac{t}{\sqrt{s}}\right\}\right|.\end{aligned}$$

Moreover, if $q_k=1$ *for all* $k=1,\dots,r$, *then (6.35) is trivially satisfied for any* $\gamma>0$ *and the left-hand side of (6.34) is equal to zero.*

Proof. To prove (i) we note that, without loss of generality, we can assume that $\|\xi\|_{l^\infty}=1$. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables with $\mathbb{P}(\delta_j=1)=\tilde{q}_j=q_k$, for $j\in\{N_{k-1}+1,\dots,N_k\}$ and $1\leq k\leq r$. A key observation that will be crucial below is that

$$\begin{aligned} P_\Delta^\perp U^*(q_1^{-1}P_{\Omega_1}\oplus\dots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta\xi &= \sum_{j=1}^N P_\Delta^\perp U^*\tilde{q}_j^{-1}\delta_j(e_j\otimes e_j)UP_\Delta\xi \\ &= \sum_{j=1}^N P_\Delta^\perp U^*(\tilde{q}_j^{-1}\delta_j-1)(e_j\otimes e_j)UP_\Delta\xi + P_\Delta^\perp U^*P_NUP_\Delta\xi. \end{aligned} \tag{6.36}$$

We will use this equation at the end of the argument, but first we will estimate the size of the individual components of $\sum_{j=1}^N P_\Delta^\perp U^*(\tilde{q}_j^{-1}\delta_j-1)(e_j\otimes e_j)UP_\Delta\xi$. To do that define, for $1\leq j\leq N$, the random variables

$$X_j^i = \langle U^*(\tilde{q}_j^{-1}\delta_j-1)(e_j\otimes e_j)UP_\Delta\xi, e_i\rangle, \qquad i\in\Delta^c$$

and the set

$$\tilde{\mathcal{G}} = \{j : \tilde{q}_j<1, j=1,\dots,N\}.$$

Observe that $X_j^i=0$ whenever $j\notin\tilde{\mathcal{G}}$.

We will show using Bernstein's inequality that, for each $i\in\Delta^c$ and $t>0$,

$$\mathbb{P}\left(\left|\sum_{j=1}^N X_j^i\right|>t\right) \leq 4\exp\left(-\frac{t^2/4}{\Upsilon+\Lambda t/3}\right). \tag{6.37}$$

To prove the claim, we need to estimate $\mathbb{E}\left(|X_j^i|^2\right)$ and $|X_j^i|$. First note that,

$$\mathbb{E}\left(|X_j^i|^2\right) = (\tilde{q}_j^{-1}-1)|\langle e_j, UP_\Delta\xi\rangle|^2|\langle e_j, Ue_i\rangle|^2$$

and recall that $|\langle e_j, Ue_i\rangle|^2\leq\mu_{N_k}$ for $j\in\{N_{k-1}+1,\dots,N_k\}$. Hence, it follows that

$$\sum_{j=1}^N \mathbb{E}\left(|X_j^i|^2\right) \leq \sum_{k=1}^r (q_k^{-1}-1)\mu_{N_{k-1}}\|P_{N_k}^{N_{k-1}}UP_\Delta\xi\|^2 \leq \sup_{\zeta\in\Theta}\left\{\sum_{k=1}^r (q_k^{-1}-1)\mu_{N_{k-1}}\|P_{N_k}^{N_{k-1}}U\zeta\|^2\right\},$$

$$\Theta = \{\zeta : \|P_{M_k}^{M_{k-1}}\zeta\|\leq\sqrt{s_k},\ 1\leq k\leq r\}.$$

Observe that, clearly, the supremum in the above bound is attained for some $\tilde{\zeta}\in\Theta$ and let $\tilde{s}_k=\|P_{N_k}^{N_{k-1}}U\tilde{\zeta}\|^2$. Hence, we have

$$\sum_{j=1}^N \mathbb{E}\left(|X_j^i|^2\right) \leq \sum_{k=1}^r (q_k^{-1}-1)\mu_{N_{k-1}}\tilde{s}_k. \tag{6.38}$$

Note that it is clear from the definition that $s_k \leq S_k(s_1, \dots, s_r)$ for $1 \leq k \leq r$. Also, observe that from the fact that $\|U\| \leq 1$ and the definition of Θ, it follows that

$$\tilde{s}_1 + \ldots + \tilde{s}_r = \sum_{k=1}^{r} \|P_{N_k}^{N_{k-1}} U P_\Delta \zeta\|^2 \leq \|U P_\Delta \zeta\|^2 = \|\zeta\|^2 \leq s_1 + \ldots + s_r.$$

To estimate $|X_j^i|$ we start by observing that, by the triangle inequality, the fact that $\|\xi\|_{l^\infty} = 1$ and Holder's inequality, it follows that

$$|\langle \xi, P_\Delta U^* e_j \rangle| \leq \sum_{k=1}^{r} |\langle P_{M_k}^{M_{k-1}} \xi, P_\Delta U^* e_j \rangle|,$$

$$|\langle P_{M_k}^{M_{k-1}} \xi, P_\Delta U^* e_j \rangle| \leq \|P_{N_l}^{N_{l-1}} U P_{\Delta_k}\|_{l^\infty \to l^\infty}, \quad j \in \{N_{l-1}+1, \dots, N_l\}, \quad l \in \{1, \dots, r\},$$

and

$$|\langle P_{M_k}^{M_{k-1}} \xi, P_\Delta U^* e_j \rangle| \leq \mu_{\mathbf{N},\mathbf{M}}(l,k) \cdot s_k.$$

Hence, it follows that for $1 \leq j \leq N$ and $i \in \Delta^c$,

$$\begin{aligned} &|X_j^i| = \tilde{q}_j^{-1} |(\delta_j - \tilde{q}_j)| |\langle \xi, P_\Delta U^* e_j \rangle| |\langle e_j, U e_i \rangle|, \\ &\leq \max_{\substack{1 \leq k \leq r \\ k \in \mathcal{G}}} \left\{ \frac{N_k - N_{k-1}}{m_k} \cdot (\min\{\mu_{\mathbf{N},\mathbf{M}}(k,1) \cdot s_1, \kappa_{\mathbf{N},\mathbf{M}}(k,1)\} + \ldots + \mu_{\mathbf{N},\mathbf{M}}(k,r) \cdot s_r, \kappa_{\mathbf{N},\mathbf{M}}(k,r)\}) \right\}. \end{aligned} \tag{6.39}$$

Now, clearly $\mathbb{E}(X_j^i) = 0$ for $1 \leq j \leq N$ and $i \in \Delta^c$. Thus, by applying Bernstein's inequality to $\mathrm{Re}(X_j^i)$ and $\mathrm{Im}(X_j^i)$ for $j = 1, \dots, N$, via (6.38) and (6.39), the claim (6.37) follows.

Now, by (6.37), (6.36) and the assumed weak Balancing property (wBP), it follows that

$$\begin{aligned} &\mathbb{P}\left(\|P_M P_\Delta^\perp U^* (q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}) U P_\Delta \xi\|_{l^\infty} > \beta\right) \\ &\leq \sum_{i \in \Delta^c \cap \{1,\dots,M\}} \mathbb{P}\left(\left| \sum_{j=1}^{N} X_j^i + \langle P_M P_\Delta^\perp U^* P_N^\perp U P_\Delta \xi, e_i \rangle \right| > \beta \right) \\ &\leq \sum_{i \in \Delta^c \cap \{1,\dots,M\}} \mathbb{P}\left(\left| \sum_{j=1}^{N} X_j^i \right| > \beta - \|P_M P_\Delta^\perp U^* P_N U P_\Delta\|_{l^\infty} \right) \\ &\leq 4(M-s) \exp\left(-\frac{t^2/4}{\Upsilon + \Lambda t/3} \right), \quad t = \frac{1}{2}\beta, \qquad \text{by (6.37), (wBP)}, \end{aligned}$$

Also,

$$4(M-s) \exp\left(-\frac{t^2/4}{\Upsilon + \Lambda t/3} \right) \leq \gamma$$

when

$$\log\left(\frac{4}{\gamma}(M-s) \right)^{-1} \geq \left(\frac{4\Upsilon}{t^2} + \frac{4\Lambda}{3t} \right).$$

And this concludes the proof of (i). To prove (ii), for $t > 0$, suppose that there is a set $\Lambda_t \subset \mathbb{N}$ such that

$$\mathbb{P}\left(\sup_{i \in \Lambda_t} |\langle P_\Delta^\perp U^* (q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}) U P_\Delta \eta, e_i \rangle| > t \right) = 0, \qquad |\Lambda_t^c| < \infty.$$

Then, as before, by (6.37), (6.36) and the assumed strong Balancing property (sBP), it follows

that

$$\begin{aligned}
&\mathbb{P}\left(\|P_\Delta^\perp U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta\xi\|_{l^\infty}>\beta\right)\\
&\leq \sum_{i\in\Delta^c\cap\Lambda_t^c}\mathbb{P}\left(\left|\sum_{j=1}^N X_j^i+\langle P_\Delta^\perp U^*P_N^\perp UP_\Delta\xi,e_i\rangle\right|>\beta\right)\\
&\leq \sum_{i\in\Delta^c\cap\Lambda_t^c}\mathbb{P}\left(\left|\sum_{j=1}^N X_j^i\right|>\beta-\|P_\Delta^\perp U^*P_NUP_\Delta\|_{l^\infty}\right)\\
&\leq 4(|\Lambda_t^c|-s)\exp\left(-\frac{t^2/4}{\Upsilon+\Lambda t/3}\right),\quad t=\frac{1}{2}\beta,\qquad \text{by (6.37), (sBP)},\\
&<\gamma
\end{aligned}$$

whenever

$$\log\left(\frac{4}{\gamma}(|\Lambda_t^c|-s)\right)^{-1}\geq\left(\frac{4\Upsilon}{t^2}+\frac{4\Lambda}{3t}\right).$$

Hence, it remains to obtain a bound on $|\Lambda_t^c|$. Let

$$\theta(q_1,\ldots,q_r,t,s)=\left\{i\in\mathbb{N}:\max_{\substack{\Gamma_1\subset\{1,\ldots,M\},\quad|\Gamma_1|=s\\ \Gamma_{2,j}\subset\{N_{j-1}+1,\ldots,N_j\},\quad j=1,\ldots,r}}\|P_{\Gamma_1}U^*(q_1^{-1}P_{\Gamma_{2,1}}\oplus\ldots\oplus q_r^{-1}P_{\Gamma_{2,r}})Ue_i\|>\frac{t}{\sqrt{s}}\right\}.$$

Clearly, $\Delta_t^c\subset\theta(q_1,\ldots,q_r,t,s)$ and

$$\|P_{\Gamma_1}U^*(q_1^{-1}P_{\Gamma_{2,1}}\oplus\ldots\oplus q_r^{-1}P_{\Gamma_{2,r}})Ue_i\|\leq\max_{1\leq j\leq r}q_j^{-1}\|P_NUP_{i-1}^\perp\|\to 0$$

as $i\to\infty$. So, $|\theta(q_1,\ldots,q_r,t,s)|<\infty$. Furthermore, since $\tilde\theta(\{q_k\}_{k=1}^r,t,\{N_k\}_{k=1}^r,s,M)$ is a decreasing function in t, for all $t\geq\frac{1}{8}$,

$$|\theta(q_1,\ldots,q_r,t,s)|<\tilde\theta(\{q_k\}_{k=1}^r,1/8,\{N_k\}_{k=1}^r,s,M)$$

thus, we have proved (ii). The statements at the end of (i) and (ii) are clear from the reasoning above. ☐

Proposition 6.12. *Consider the same setup as in Proposition 6.11. If N and K satisfy the weak Balancing Property with respect to U, M and s, then, for $\xi\in\mathcal{H}$ and $t,\gamma>0$, we have*

$$\mathbb{P}(\|P_\Delta U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta-P_\Delta)\xi\|_{l^\infty}>\tilde\alpha\|\xi\|_{l^\infty})\leq\gamma, \tag{6.40}$$

$$\tilde\alpha=\left(2\log_2^{1/2}\left(4\sqrt{s}KM\right)\right)^{-1},$$

provided that

$$\begin{aligned}
1&\gtrsim\Lambda\cdot\left(\log\left(s\gamma^{-1}\right)+1\right)\cdot\log\left(\sqrt{s}KM\right),\\
1&\gtrsim\Upsilon\cdot\left(\log\left(s\gamma^{-1}\right)+1\right)\cdot\log\left(\sqrt{s}KM\right),
\end{aligned}$$

where Λ and Υ are defined in (6.32) and (6.33). Also,

$$\mathbb{P}(\|P_\Delta U^*(q_1^{-1}P_{\Omega_1}\oplus\ldots\oplus q_r^{-1}P_{\Omega_r})UP_\Delta-P_\Delta)\xi\|_{l^\infty}>\frac{1}{2}\|\xi\|_{l^\infty})\leq\gamma \tag{6.41}$$

provided that

$$1\gtrsim\Lambda\cdot\left(\log\left(s\gamma^{-1}\right)+1\right),\quad 1\gtrsim\Upsilon\cdot\left(\log\left(s\gamma^{-1}\right)+1\right).$$

Moreover, if $q_k=1$ for all $k=1,\ldots,r$, then the left-hand sides of (6.40) and (6.41) are equal to zero.

Proof. Without loss of generality we may assume that $\|\xi\|_{l^\infty} = 1$. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables with $\mathbb{P}(\delta_j = 1) = q_k$, with $j \in \{N_{k-1}+1, \dots, N_k\}$ and $1 \le k \le r$. Let also, for $j \in \mathbb{N}$, $\eta_j = (UP_\Delta)^* e_j$. Then, after observing that

$$P_\Delta U^*(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}) U P_\Delta = \sum_{j=1}^N q_j^{-1} \delta_j \eta_j \otimes \bar{\eta}_j, \quad P_\Delta U^* P_N U P_\Delta = \sum_{j=1}^N \eta_j \otimes \bar{\eta}_j,$$

it follows immediately that

$$P_\Delta U^*(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}) U P_\Delta - P_\Delta = \sum_{j=1}^N (q_j^{-1}\delta_j - 1)\eta_j \otimes \bar{\eta}_j - (P_\Delta U^* P_N U P_\Delta - P_\Delta). \tag{6.42}$$

As in the proof of Proposition 6.11 Our goal is to eventually use Bernstein's inequality and the following is therefore a setup for that. Define, for $1 \le j \le N$, the random variables

$$Z_j^i = \langle (q_j^{-1}\delta_j - 1)(\eta_j \otimes \bar{\eta}_j)\xi, e_i \rangle, \qquad i \in \Delta.$$

We claim that, for $t > 0$,

$$\mathbb{P}\left(\left| \sum_{j=1}^N Z_j^i \right| > t \right) \le 4 \exp\left(- \frac{t^2/4}{\Upsilon + \Lambda t/3} \right), \qquad i \in \Delta. \tag{6.43}$$

Now, clearly $\mathbb{E}(Z_j^i) = 0$, so we may use Bernstein's inequality. Thus, we need to estimate $\mathbb{E}\left(|Z_j^i|^2\right)$ and $|Z_j^i|$. We will start with $\mathbb{E}\left(|Z_j^i|^2\right)$. Note that

$$\mathbb{E}\left(|Z_j^i|^2\right) = (q_j^{-1} - 1)|\langle e_j, UP_\Delta \xi\rangle|^2 |\langle e_j, U e_i \rangle|^2. \tag{6.44}$$

Thus, we can argue exactly as in the proof of Proposition 6.11 and deduce that

$$\sum_{j=1}^N \mathbb{E}\left(|Z_j^i|^2\right) \le \sum_{k=1}^r (q_k^{-1} - 1)\mu_{N_{k-1}} \tilde{s}_k, \tag{6.45}$$

where $s_k \le S_k(s_1, \dots, s_r)$ for $1 \le k \le r$ and $\tilde{s}_1 + \ldots + \tilde{s}_r \le s_1 + \ldots + s_r$.

To estimate $|Z_j^i|$ we argue as in the proof of Proposition 6.11 and obtain

$$|Z_j^i| \le \max_{1 \le k \le r} \left\{ \frac{N_k - N_{k-1}}{m_k} \cdot (\min\{\mu_{\mathbf{N},\mathbf{M}}(k,1) \cdot s_1, \kappa_{\mathbf{N},\mathbf{M}}(k,1)\} + \ldots + \mu_{\mathbf{N},\mathbf{M}}(k,r) \cdot s_r, \kappa_{\mathbf{N},\mathbf{M}}(k,r)\}) \right\}. \tag{6.46}$$

Thus, by applying Bernstein's inequality to $\mathrm{Re}(Z_1^i), \dots, \mathrm{Re}(Z_N^i)$ and $\mathrm{Im}(Z_1^i), \dots, \mathrm{Im}(Z_N^i)$ we obtain, via (6.45) and (6.46) the estimate (6.43), and we have proved the claim.

Now armed with (6.43) we can deduce that , by (6.36) and the assumed weak Balancing property (wBP), it follows that

$$\begin{aligned}
&\mathbb{P}\left(\|P_\Delta U^*(q_1^{-1} P_{\Omega_1} \oplus \ldots \oplus q_r^{-1} P_{\Omega_r}) U P_\Delta - P_\Delta)\xi\|_{l^\infty} > \tilde{\alpha} \right) \\
&\le \sum_{i \in \Delta} \mathbb{P}\left(\left| \sum_{j=1}^N Z_j^i + \langle (P_\Delta U^* P_N U P_\Delta - P_\Delta)\xi, e_i \rangle \right| > \tilde{\alpha} \right) \\
&\le \sum_{i \in \Delta} \mathbb{P}\left(\left| \sum_{j=1}^N Z_j^i \right| > \tilde{\alpha} - \|P_M U^* P_N U P_M - P_M\|_{l^1} \right), \\
&\le 4\,s \exp\left(- \frac{t^2/4}{\Upsilon + \Lambda t/3} \right), \quad t = \tilde{\alpha}, \qquad \text{by (6.43), (wBP)}.
\end{aligned} \tag{6.47}$$

Also,

$$4s \exp\left(- \frac{t^2/4}{\Upsilon + \Lambda t/3} \right) \le \gamma, \tag{6.48}$$

when

$$1 \geq \left(\frac{4\Upsilon}{t^2} + \frac{4}{3t}\Lambda \right) \cdot \log\left(\frac{4s}{\gamma} \right).$$

And this gives the first part of the proposition. Also, the fact that the left hand side of (6.40) is zero when $q_k = 1$ for $1 \leq k \leq r$ is clear from (6.48). Note that (ii) follows by arguing exactly as above and replacing $\tilde{\alpha}$ by $\frac{1}{4}$.

□

6.3 Proofs of Propositions 6.3 and 6.4

The proof of the propositions relies on an idea that originated in a paper by D. Gross [28], namely, the golfing scheme. The variant we are using here is based on an idea from [2]. However, the informed reader will recognise that the setup here differs substantially from both [28] and [2]. See also [12] for other examples of the use of the golfing scheme.

Proof of Proposition 6.3. We start by mentioning that converting from the Bernoulli sampling model and uniform sampling model has become standard in the literature. In particular, one can do this by showing that the Bernoulli model implies (up to a constant) the uniform sampling model in each of the conditions in Proposition 6.1. This is straightforward and the reader may consult [14, 13, 27] for details. We will therefore consider (without loss of generality) only the multi-level Bernoulli sampling scheme.

Recall that we are using the following Bernoulli sampling model: Given $N_0 = 0, N_1, \ldots, N_r \in \mathbb{N}$ we let

$$\{N_{k-1}+1, \ldots, N_k\} \supset \Omega_k \sim \mathrm{Ber}\left(q_k\right), \quad q_k = \frac{m_k}{N_k - N_{k-1}}.$$

Note that we may replace this Bernoulli sampling model with the following equivalent sampling model (see [2]):

$$\Omega_k = \Omega_k^1 \cup \Omega_k^2 \cup \cdots \cup \Omega_k^u, \qquad \Omega_k^j \sim \mathrm{Ber}(q_k^j), \qquad 1 \leq k \leq r,$$

for some $u \in \mathbb{N}$ with

$$(1-q_k^1)(1-q_k^2)\cdots(1-q_k^u) = (1-q_k). \tag{6.49}$$

The latter model is the one we will use throughout the proof and the specific value of u will be chosen later. Note also that because of overlaps we will have

$$q_k^1 + q_k^2 + \ldots + q_k^u \geq q_k, \qquad 1 \leq k \leq r. \tag{6.50}$$

The strategy of the proof is to show the existence of a $\rho \in \mathrm{ran}(U^*(P_{\Omega_1} \oplus \ldots \oplus P_{\Omega_r}))$ that satisfies (i), (i*)-(iv) in Proposition 6.1 with probability exceeding $1-\epsilon$.

Step I: The construction of ρ: We start by defining $\gamma = \epsilon/6$ (the reason for this particular choice will become clear later). We also define a number of quantities (and the reason for these choices will become clear later in the proof):

$$u = 4\lceil \log(\gamma^{-1}) + v \rceil, \qquad v = \lceil \log_2(8KM\sqrt{s}) \rceil, \tag{6.51}$$

as well as

$$\{q_k^i : 1 \leq k \leq r, 1 \leq i \leq u\}, \quad \{\alpha_i\}_{i=1}^u, \quad \{\beta_i\}_{i=1}^u$$

by

$$q_k^1 = q_k^2 = \frac{1}{4}q_k, \qquad \tilde{q}_k = q_k^3 = \ldots = q_k^u, \qquad q_k = (N_k - N_{k-1})m_k^{-1}, \quad 1 \leq k \leq r, \tag{6.52}$$

with

$$(1-q_k^1)(1-q_k^2)\cdots(1-q_k^u) = (1-q_k)$$

and

$$\alpha_1 = \alpha_2 = (2\log_2^{1/2}(4KM\sqrt{s}))^{-1}, \qquad \alpha_i = 1/2, \quad 3 \leq i \leq u, \tag{6.53}$$

as well as

$$\beta_1 = \beta_2 = \frac{1}{4}, \qquad \beta_i = \frac{1}{4}\log_2(4KM\sqrt{s}), \quad 3 \le i \le u. \tag{6.54}$$

Consider now the following construction of ρ. We will define recursively the sequences $\{Z_i\}_{i=0}^u \subset \mathcal{H}$, $\{Y_i\}_{i=1}^u \subset \mathcal{H}$ and $\{\omega_i\}_{i=0}^u \subset \mathbb{N}$ as follows: first let $\omega_0 = \{0\}$, $\omega_1 = \{0,1\}$ and $\omega_2 = \{0,1,2\}$. Then define recursively, for $i \ge 3$, the following:

$$\omega_i = \begin{cases} \omega_{i-1} \cup \{i\} & \text{if } \|(P_\Delta - P_\Delta U^*(\frac{1}{q_1^i}P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta)Z_{i-1}\|_{l^\infty} \le \alpha_i \|P_{\Delta_k} Z_{i-1}\|_{l^\infty}, \\ & \text{and } \|P_M P_\Delta^\perp U^*(\frac{1}{q_1^i}P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta Z_{i-1}\|_{l^\infty} \le \beta_i \|Z_{i-1}\|_{l^\infty}, \\ \omega_{i-1} & \text{otherwise,} \end{cases}$$

$$Y_i = \begin{cases} \sum_{j\in\omega_i} U^*(\frac{1}{q_1^j}P_{\Omega_1^j} \oplus \ldots \oplus \frac{1}{q_r^j}P_{\Omega_r^j})UZ_{j-1} & \text{if } i \in \omega_i, \\ Y_{i-1} & \text{otherwise,} \end{cases} \qquad i \ge 1, \tag{6.55}$$

$$Z_i = \begin{cases} \mathrm{sgn}(x_0) - P_\Delta Y_i & \text{if } i \in \omega_i, \\ Z_{i-1} & \text{otherwise,} \end{cases} \qquad i \ge 1, \qquad Z_0 = \mathrm{sgn}(x_0).$$

Now, let $\{A_i\}_{i=1}^2$ and $\{B_i\}_{i=1}^5$ denote the following events

$$\begin{aligned} A_i: &\quad \|(P_\Delta - U^*(\frac{1}{q_1^i}P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta)Z_{i-1}\|_{l^\infty} \le \alpha_i \,\|Z_{i-1}\|_{l^\infty}, \qquad i = 1,2, \\ B_i: &\quad \|P_M P_\Delta^\perp U^*(\frac{1}{q_1^i}P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i}P_{\Omega_r^i})UP_\Delta Z_{i-1}\|_{l^\infty} \le \beta_i \|Z_{i-1}\|_{l^\infty}, \qquad i = 1,2, \\ B_3: &\quad \|P_\Delta U^*(\frac{1}{q_1}P_{\Omega_1} \oplus \ldots \oplus \frac{1}{q_r}P_{\Omega_r})UP_\Delta - P_\Delta\| \le 1/4, \\ B_4: &\quad |\omega_u| \ge v, \\ B_5: &\quad (\cap_{i=1}^2 A_i) \cap (\cap_{i=1}^4 B_i). \end{aligned} \tag{6.56}$$

Also, let $\tau(j)$ denote the j^{th} element in ω_u (e.g. $\tau(0) = 0, \tau(1) = 1, \tau(2) = 2$ etc.) and finally define ρ by

$$\rho = \begin{cases} Y_{\tau(v)} & \text{if } B_5 \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Note that, clearly, $\rho \in \mathrm{ran}(U^*P_\Omega)$, and we just need to show that when the event B_5 occurs, then (i), (i*)-(iv) in Proposition 6.1 will follow.

Step II: $B_5 \Rightarrow$ **(i)**, **(i*)**. To see that the assertion is true, note that if B_5 occurs then B_3 occurs, which immediately implies that $\|(P_\Delta U^*\left(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r}\right)UP_\Delta|_{P_\Delta(\mathcal{H})})^{-1}\| \le \frac{4}{3}$, yielding (i). As for (i*), note that if B_3 occurs, then it follows that

$$\|P_\Delta U^*\left(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r}\right)UP_\Delta\| \le \frac{5}{4}. \tag{6.57}$$

Also, observe that

$$\begin{aligned} &\|P_\Delta U^*\left(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r}\right)\|^2 = \|\left(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r}\right)UP_\Delta\|^2 \\ &= \sup_{\|\eta\|=1} \|\left(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r}\right)UP_\Delta\eta\|^2 \\ &= \sup_{\|\eta\|=1} \sum_{k=1}^r \|q_k^{-1}P_{\Omega_k}UP_\Delta\eta\|^2 \le \frac{1}{q}\sup_{\|\eta\|=1}\sum_{k=1}^r q_k^{-1}\|P_{\Omega_k}UP_\Delta\eta\|^2, \quad \frac{1}{q} = \max_{1\le k\le r}\{\frac{1}{q_k}\} \\ &= \frac{1}{q}\sup_{\|\eta\|=1}\langle P_\Delta U^*\left(\sum_{k=1}^r q_k^{-1}P_{\Omega_k}\right)UP_\Delta\eta, \eta\rangle \le \frac{1}{q}\|P_\Delta U^*\left(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r}\right)UP_\Delta\|. \end{aligned} \tag{6.58}$$

Thus, (6.57) and (6.58) imply $\|P_\Delta U^*\left(q_1^{-1}P_{\Omega_1} \oplus \ldots \oplus q_r^{-1}P_{\Omega_r}\right)\| \le \sqrt{\frac{5}{4q}}$, and we have shown (i*).

Step III: $B_5 \Rightarrow$ **(ii)**, **(iii)**. To show the assertion, we start by making the following observations: By the construction of $Z_{\tau(i)}$ and the fact that $Z_0 = \mathrm{sgn}(x_0)$, it follows that

$$\begin{aligned}
Z_{\tau(i)} &= Z_0 - (P_\Delta U^*(\frac{1}{q_1^{\tau(1)}} P_{\Omega_1^{\tau(1)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(1)}} P_{\Omega_r^{\tau(i)}}) U P_\Delta) Z_0 \\
&\quad + \ldots + P_\Delta U^*(\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}}) U P_\Delta) Z_{\tau(i-1)}) \\
&= Z_{\tau(i-1)} - P_\Delta U^*(\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}}) U P_\Delta) Z_{\tau(i-1)} \qquad i \leq |\omega_u|,
\end{aligned}$$

so we immediately get that

$$Z_{\tau(i)} = (P_\Delta - P_\Delta U^*(\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}}) U P_\Delta) Z_{\tau(i-1)}, \qquad i \leq |\omega_u|.$$

Hence, if the event B_5 occurs, we have, by the choices in (6.53) and (6.54)

$$\|\rho - \mathrm{sgn}(x_0)\| = \|Z_{\tau(v)}\| \leq \sqrt{s}\|Z_{\tau(v)}\|_{l^\infty} \leq \sqrt{s}\prod_{i=1}^{v} \alpha_{\tau(i)} \leq \frac{\sqrt{s}}{2^v} \leq \frac{1}{8K}, \tag{6.59}$$

since we have chosen $v = \lceil \log_2(8KM\sqrt{s}) \rceil$. Also,

$$\begin{aligned}
\|P_M P_\Delta^\perp \rho\|_{l^\infty} &\leq \sum_{i=1}^{v} \|P_M P_\Delta^\perp U^*(\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}}) U P_\Delta Z_{\tau(i-1)}\|_{l^\infty} \\
&\leq \sum_{i=1}^{v} \beta_{\tau(i)} \|Z_{\tau(i-1)}\|_{l^\infty} \leq \sum_{i=1}^{v} \beta_{\tau(i)} \prod_{j=1}^{i-1} \alpha_{\tau(j)} \\
&\leq \frac{1}{4}(1 + \frac{1}{2\log_2^{1/2}(a)} + \frac{\log_2(a)}{2^3 \log_2(a)} + \ldots + \frac{1}{2^{v-1}}) \leq \frac{1}{2}, \qquad a = 8K\sqrt{s}.
\end{aligned} \tag{6.60}$$

In particular, (6.59) and (6.60) imply (ii) and (iii) in Proposition 6.1.

Step IV: $B_5 \Rightarrow$ **(iv)**. To show that, note that we may write the already constructed ρ as $\rho = U^* P_\Omega w$ where

$$w = \sum_{i=1}^{v} w_i, \quad w_i = \left(\frac{1}{q_1^{\tau(i)}} P_{\Omega_1} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r} \right) U P_\Delta Z_{\tau(i-1)}.$$

To estimate $\|w\|$ we simply compute

$$\begin{aligned}
\|w_i\|^2 &= \left\langle \left(\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}} \right) U P_\Delta Z_{\tau(i-1)}, \left(\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}} \right) U P_\Delta Z_{\tau(i-1)} \right\rangle \\
&= \sum_{k=1}^{r} \left(\frac{1}{q_k^{\tau(i)}} \right)^2 \|P_{\Omega_k^{\tau(i)}} U Z_{\tau(i-1)}\|^2,
\end{aligned}$$

and then use the assumption that the event B_5 holds to deduce that

$$\begin{aligned}
&\sum_{k=1}^{r} \left(\frac{1}{q_k^{\tau(i)}} \right)^2 \|P_{\Omega_k^{\tau(i)}} U Z_{\tau(i-1)}\|^2 \leq \max_{1 \leq k \leq r} \left\{ \frac{1}{q_k^{\tau(i)}} \right\} \langle \sum_{k=1}^{r} \frac{1}{q_k^{\tau(i)}} P_\Delta U^* P_{\Omega_k^{\tau(i)}} U Z_{\tau(i-1)}, Z_{\tau(i-1)} \rangle \\
&= \max_{1 \leq k \leq r} \left\{ \frac{1}{q_k^{\tau(i)}} \right\} \langle \left(\sum_{k=1}^{r} \frac{1}{q_k^{\tau(i)}} P_\Delta U^* P_{\Omega_k^{\tau(i)}} U - P_\Delta \right) Z_{\tau(i-1)}, Z_{\tau(i-1)} \rangle + \|Z_{\tau(i-1)}\|^2 \\
&\leq \max_{1 \leq k \leq r} \left\{ \frac{1}{q_k^{\tau(i)}} \right\} \left(\|Z_{\tau(i-1)}\| \|Z_{\tau(i)}\| + \|Z_{\tau(i-1)}\|^2 \right) \\
&\leq \max_{1 \leq k \leq r} \left\{ \frac{1}{q_k^{\tau(i)}} \right\} s \left(\|Z_{\tau(i-1)}\|_{l^\infty} \|Z_{\tau(i)}\|_{l^\infty} + \|Z_{\tau(i-1)}\|_{l^\infty}^2 \right) \leq \max_{1 \leq k \leq r} \left\{ \frac{1}{q_k^{\tau(i)}} \right\} s(\alpha_i + 1) \left(\prod_{j=1}^{i-1} \alpha_j \right)^2,
\end{aligned}$$

where the last inequality follows from the assumption that the event B_5 holds. Hence

$$\|w\| \leq \sqrt{s} \sum_{i=1}^{v} \left(\max_{1 \leq k \leq r} \left\{ \frac{1}{\sqrt{q_k^{\tau(i)}}} \right\} \sqrt{\alpha_i + 1} \prod_{j=1}^{i-1} \alpha_j \right) \tag{6.61}$$

Note that, due to the fact that $q_k^1 + \ldots + q_k^u \geq q_k$, we have that

$$\tilde{q}_k \geq \frac{m_k}{4(N_k - N_{k-1})} \frac{1}{4\left\lceil \log(\gamma^{-1}) + \lceil \log_2(8KM\sqrt{s}) \rceil \right\rceil - 2}.$$

This gives, in combination with the chosen values of $\{\alpha_j\}$ and (6.61) that

$$\begin{aligned}
\|w\| &\leq 2\sqrt{s} \max_{1\leq k \leq r} \sqrt{\frac{N_k - N_{k-1}}{m_k}} \left(1 + \frac{1}{2\log_2^{1/2}(4KM\sqrt{s})}\right)^{3/2} \\
&+ \sqrt{s} \max_{1\leq k \leq r} \sqrt{\frac{N_k - N_{k-1}}{m_k}} \cdot \sqrt{\frac{3}{2}} \cdot \frac{\sqrt{\left\lceil \log(\gamma^{-1}) + \lceil \log_2(8KM\sqrt{s}) \rceil \right\rceil - 2}}{\log_2(4KM\sqrt{s})} \cdot \sum_{i=3}^{v} \frac{1}{2^{i-3}} \\
&\leq 2\sqrt{s} \max_{1\leq k \leq r} \sqrt{\frac{N_k - N_{k-1}}{m_k}} \left(\left(\frac{3}{2}\right)^{3/2} + \left(\frac{3}{2}\right)^{1/2} \sqrt{1 + \frac{\log_2(\gamma^{-1}) + 1}{\log_2(4KM\sqrt{s})}} \right) \\
&\leq \sqrt{6s} \max_{1\leq k \leq r} \sqrt{\frac{N_k - N_{k-1}}{m_k}} \left(\frac{3}{2} + \sqrt{1 + \frac{\log_2(\gamma^{-1}) + 1}{\log_2(4KM\sqrt{s})}} \right).
\end{aligned} \tag{6.62}$$

Step V: The weak balancing property, (6.10) and (6.11) $\Rightarrow \mathbb{P}(A_1^c \cup A_2^c \cup B_1^c \cup B_2^c \cup B_3^c) \leq 5\gamma$.

To see this, note that by Proposition 6.12 we immediately get that $\mathbb{P}(A_1^c) \leq \gamma$ and $\mathbb{P}(A_2^c) \leq \gamma$ as long as the weak balancing property and

$$\begin{aligned}
1 &\gtrsim \Lambda \cdot \left(\log\left(s\gamma^{-1}\right) + 1\right) \cdot \log\left(\sqrt{s}KM\right), \\
1 &\gtrsim \Upsilon \cdot \left(\log\left(s\gamma^{-1}\right) + 1\right) \cdot \log\left(\sqrt{s}KM\right),
\end{aligned} \tag{6.63}$$

are satisfied, where $K = \max_{1\leq k \leq r}(N_k - N_{k-1})/m_k$

$$\Lambda = \max_{1\leq k \leq r} \left\{ \frac{N_k - N_{k-1}}{m_k} \cdot (\nu_{k,1} + \ldots + \nu_{k,r}) \right\}, \qquad \nu_{k,l} = \min(\mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot s_l, \kappa_{\mathbf{N},\mathbf{M}}(k,l)), \tag{6.64}$$

and

$$\Upsilon = \left(\frac{N_1 - N_0}{m_1} - 1\right) \cdot \mu_{N_0} \cdot \tilde{s}_1 + \ldots + \left(\frac{N_r - N_{r-1}}{m_r} - 1\right) \cdot \mu_{N_{r-1}} \cdot \tilde{s}_r, \tag{6.65}$$

where $\tilde{s}_1 + \ldots + \tilde{s}_r \leq s_1 + \ldots + s_r$ and $\tilde{s}_k \leq S_k(s_1, \ldots, s_r)$. However, clearly, (6.10) and (6.11) imply (6.63). Also, Proposition 6.11 yields that $\mathbb{P}(B_1^c) \leq \gamma$ and $\mathbb{P}(B_2^c) \leq \gamma$ as long as the weak balancing property and

$$1 \gtrsim \Lambda \cdot \log\left(\frac{4}{\gamma}(M - s)\right), \qquad 1 \gtrsim \Upsilon \cdot \log\left(\frac{4}{\gamma}(M - s)\right) \tag{6.66}$$

are satisfied. However, again, (6.10) and (6.11) imply (6.66). Finally, to bound $\mathbb{P}(B_3^c)$, we use Proposition 6.8 to deduce that $\mathbb{P}(B_3^c) \leq \gamma$ when the weak balancing property and

$$1 \geq \Lambda \cdot \left(\log\left(\gamma^{-1} s\right) + 1\right) \tag{6.67}$$

is satisfied. But (6.10) implies (6.67), and we are done.

Step VI: The weak balancing property, (6.10) and (6.11) $\Rightarrow \mathbb{P}(B_4^c) \leq \gamma$.

To see this, define the random variables $X_1, \ldots X_{u-2}$ by

$$X_j = \begin{cases} 0 & \omega_{j+2} \neq \omega_{j+1}, \\ 1 & w_{j+2} = w_{j+1}. \end{cases} \tag{6.68}$$

We immediately observe that

$$\mathbb{P}(B_4^c) = \mathbb{P}(|\omega_u| < v) = \mathbb{P}(X_1 + \ldots + X_{u-2} > u - v). \tag{6.69}$$

Note that if we can provide a bound

$$\frac{1}{2} \geq \mathbb{P}(X_j = 1), \qquad j = 1, \ldots, u-2, \tag{6.70}$$

then, by the standard Chernoff bound ([36, Theorem 2.1]), it follows that, for $t > 0$,

$$\mathbb{P}\left(X_1 + \ldots + X_{u-2} \geq (u-2)(t + 1/2)\right) \leq e^{-2(u-2)t^2}. \tag{6.71}$$

Hence, if we let $t = (u-v)/(u-2) - 1/2$ and plug in the values of u and v from (6.51) into (6.71), then we get that

$$\mathbb{P}(B_4^c) = \mathbb{P}(X_1 + \ldots + X_{u-2} > u - v) \leq \gamma.$$

Thus, the proposition follows from the next claim.

Claim: The weak balancing property, (6.10) and (6.11) $\Rightarrow$ (6.70). To prove the claim we first observe that $X_j = 0$ when

$$\|(P_\Delta - P_\Delta U^*(\frac{1}{q_1^i} P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i} P_{\Omega_r^i}) U P_\Delta) Z_{i-1}\|_{l^\infty} \leq \frac{1}{2} \|Z_{i-1}\|_{l^\infty}$$

$$\|P_M P_\Delta^\perp U^*(\frac{1}{q_1^i} P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i} P_{\Omega_r^i}) U P_\Delta Z_{i-1}\|_{l^\infty} \leq \frac{1}{4} \log_2(4KM\sqrt{s}) \|Z_{i-1}\|_{l^\infty}, \qquad i = j+2,$$

where we recall from (6.52) that

$$q_k^3 = q_k^4 = \ldots = q_k^u = \tilde{q}_k, \qquad 1 \leq k \leq r.$$

Thus, by choosing $\tilde{\gamma} = 1/4$ in (6.41) in Proposition 6.12 and $\tilde{\gamma} = 1/4$ in (i) in Proposition 6.11, it follows that $\frac{1}{2} \geq \mathbb{P}(X_j = 1)$, for $j = 1, \ldots, u-2$, when the weak balancing property is satisfied and

$$(\log(4) + 1)^{-1} \geq C_1 \cdot \tilde{q}_k^{-1} \cdot (\nu_{k,1} + \ldots + \nu_{k,r}), \quad 1 \leq k \leq r \tag{6.72}$$

$$(\log(4) + 1)^{-1} \geq C_1 \cdot \left(\left(\tilde{q}_1^{-1} - 1\right) \cdot \mu_{N_1}^2 \cdot \tilde{s}_1 + \ldots + \left(\tilde{q}_r^{-1} - 1\right) \cdot \mu_{N_r}^2 \cdot \tilde{s}_r\right), \tag{6.73}$$

$$\nu_{k,l} = \min(\mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot s_l, \kappa_{\mathbf{N},\mathbf{M}}(k,l))$$

as well as

$$\frac{\log_2(4KM\sqrt{s})}{\log(16(M-s))} \geq C_2 \cdot \tilde{q}_k^{-1} \cdot (\nu_{k,1} + \ldots + \nu_{k,r}), \quad 1 \leq k \leq r \tag{6.74}$$

$$\frac{\log_2(4KM\sqrt{s})}{\log(16(M-s))} \geq C_2 \cdot \left(\left(\tilde{q}_1^{-1} - 1\right) \cdot \mu_{N_0} \cdot \tilde{s}_1 + \ldots + \left(\tilde{q}_r^{-1} - 1\right) \cdot \mu_{N_{r-1}} \cdot \tilde{s}_r\right), \tag{6.75}$$

with $K = \max_{1 \leq k \leq r}(N_k - N_{k-1})/m_k$. for some constants C_1 and C_2. Thus, to prove the claim we must demonstrate that (6.10) and (6.11) $\Rightarrow$ (6.72), (6.73), (6.74) and (6.75). We split this into two cases:

Case 1: (6.11) $\Rightarrow$ (6.75) and (6.73).

To show the assertion we must demonstrate that if, for $1 \leq k \leq r$,

$$m_k \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \hat{m}_k \cdot \log\left(KM\sqrt{s}\right), \tag{6.76}$$

where $\hat{m}_k$ satisfies

$$1 \gtrsim \left(\frac{N_1 - N_0}{\hat{m}_1} - 1\right) \cdot \mu N_0 \cdot \tilde{s}_1 + \ldots + \left(\frac{N_r - N_{r-1}}{\hat{m}_r} - 1\right) \cdot \mu_{N_{r-1}} \cdot \tilde{s}_r, \tag{6.77}$$

we get (6.75) and (6.73). To see this, note that by (6.50) we have that

$$q_k^1 + q_k^2 + (u-2)\tilde{q}_k \geq q_k, \qquad 1 \leq k \leq r, \tag{6.78}$$

so since $q_k^1 = q_k^2 = \frac{1}{4} q_k$, and by (6.78), (6.76) and the choice of μ in (6.51), it follows that

$$
\begin{aligned}
2(4(\lceil \log(\gamma^{-1}) + \lceil \log_2(8KM\sqrt{s}) \rceil \rceil) - 2)\tilde{q}_k \geq q_k &= \frac{m_k}{N_k - N_{k-1}} \\
&\geq C \frac{\hat{m}_k}{N_k - N_{k-1}} \log(s\epsilon^{-1} + 1) \log\left(KM\sqrt{s}\right),
\end{aligned}
$$

for some constant C. And this gives (by recalling that $\gamma = \epsilon/6$) that

$$
\tilde{q}_k \geq \tilde{C} \frac{\hat{m}_k}{N_k - N_{k-1}} \frac{\log(s\epsilon^{-1} + 1) \log\left(KM\sqrt{s}\right)}{\lceil \log(6\epsilon^{-1}) + \lceil \log_2(8KM\sqrt{s}) \rceil \rceil},
$$

for some constant $\tilde{C}$. Hence, given that $\tilde{C}$ is chosen appropriately, we get that $\tilde{q}_k \geq \frac{\hat{m}_k}{N_k - N_{k-1}}$, for $1 \leq k \leq r$, which, since $\hat{m}_k$ satisfies (6.77), this implies (6.75), given the appropriate choice of the constants. Note that similar reasoning also immediately implies (6.73).

Case 2: (6.10) $\Rightarrow$ (6.74) and (6.72).

To show the assertion we must demonstrate that if, for $1 \leq k \leq r$,

$$
1 \gtrsim (\log(s\epsilon^{-1}) + 1) \cdot \frac{N_k - N_{k-1}}{m_k} \cdot (\nu_{k,1} + \ldots + \nu_{k,r}) \cdot \log\left(KM\sqrt{s}\right), \tag{6.79}
$$

we obtain (6.74) and (6.72). To see this, note that by arguing as above via the fact that $q_k^1 = q_k^2 = \frac{1}{4} q_k$, and by (6.78), (6.79) and the choice of μ in (6.51) we have that

$$
\begin{aligned}
2(4(\lceil \log(\gamma^{-1}) + \lceil \log_2(8KM\sqrt{s}) \rceil \rceil) - 2)\tilde{q}_k \geq q_k &= \frac{m_k}{N_k - N_{k-1}} \\
&\geq C(\log(s\epsilon^{-1}) + 1) \cdot (\nu_{k,1} + \ldots + \nu_{k,r}) \cdot \log\left(KM\sqrt{s}\right),
\end{aligned}
$$

for some constant C. We can now argue as in Case 1 and deduce that

$$
\tilde{q}_k \geq \hat{C} \cdot (\nu_{k,1} + \ldots + \nu_{k,r})
$$

given that $\hat{C}$ is a appropriately chosen. However, that latter equation implies (6.74) and (6.72), given an appropriate choice of the constants.

This yields the last puzzle of the proof, and we are done.

□

Proof of Proposition 6.4. The proof is very close to the proof of Proposition 6.3 and we will simply point out the differences. The strategy of the proof is to show the existence of a $\rho \in \mathrm{ran}(U^*(P_{\Omega_1} \oplus \ldots \oplus P_{\Omega_r}))$ that satisfies (i), (i*)-(iv) in Proposition 6.1 with probability exceeding $1 - \epsilon$.

Step I: The construction of ρ: The construction is almost identical to the construction in the proof of Proposition 6.3, except that

$$
u = 4\lceil \log(\gamma^{-1}) + v \rceil, \qquad v = \lceil \log_2(8K\tilde{M}\sqrt{s}) \rceil, \tag{6.80}
$$

$$
\alpha_1 = \alpha_2 = (2\log_2^{1/2}(4K\tilde{M}\sqrt{s}))^{-1}, \qquad \alpha_i = 1/2, \quad 3 \leq i \leq u,
$$

as well as

$$
\beta_1 = \beta_2 = \frac{1}{4}, \qquad \beta_i = \frac{1}{4}\log_2(4K\tilde{M}\sqrt{s}), \quad 3 \leq i \leq u,
$$

and (6.55) gets changed to

$$
\omega_i = \begin{cases} \omega_{i-1} \cup \{i\} & \text{if } \|(P_\Delta - P_\Delta U^*(\frac{1}{q_1^i} P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i} P_{\Omega_r^i}) U P_\Delta) Z_{i-1}\|_{l^\infty} \leq \alpha_i \|P_{\Delta_k} Z_{i-1}\|_{l^\infty}, \\ & \text{and } \|P_\Delta^\perp U^*(\frac{1}{q_1^i} P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i} P_{\Omega_r^i}) U P_\Delta Z_{i-1}\|_{l^\infty} \leq \beta_i \|Z_{i-1}\|_{l^\infty}, \\ \omega_{i-1} & \text{otherwise,} \end{cases}
$$

and the events B_i, $i = 1, 2$ in (6.56) get replaced by

$$
\widetilde{B}_i : \qquad \|P_\Delta^\perp U^*(\frac{1}{q_1^i} P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i} P_{\Omega_r^i}) U P_\Delta Z_{i-1}\|_{l^\infty} \leq \beta_i \|Z_{i-1}\|_{l^\infty}, \qquad i = 1, 2.
$$

Step II: $B_5 \Rightarrow$ **(i)**, **(i*)**. This step is identical to Step II in the proof of Proposition 6.3.

Step III: $B_5 \Rightarrow$ **(ii)**, **(iii)**. Equation (6.60) gets changed to

$$\begin{aligned}\|P_\Delta^\perp \rho\|_{l^\infty} &\leq \sum_{i=1}^{v} \|P_\Delta^\perp U^*(\frac{1}{q_1^{\tau(i)}} P_{\Omega_1^{\tau(i)}} \oplus \ldots \oplus \frac{1}{q_r^{\tau(i)}} P_{\Omega_r^{\tau(i)}}) U P_\Delta Z_{\tau(i-1)}\|_{l^\infty} \\ &\leq \sum_{i=1}^{v} \beta_{\tau(i)} \|Z_{\tau(i-1)}\|_{l^\infty} \leq \sum_{i=1}^{v} \beta_{\tau(i)} \prod_{j=1}^{i-1} \alpha_{\tau(j)} \\ &\leq \frac{1}{4}(1 + \frac{1}{2\log_2^{1/2}(a)} + \frac{\log_2(a)}{2^3 \log_2(a)} + \ldots + \frac{1}{2^{v-1}}) \leq \frac{1}{2}, \qquad a = 8\tilde{M}K\sqrt{s}.\end{aligned}$$

Step IV: $B_5 \Rightarrow$ **(iv)**. This step is identical to Step IV in the proof of Proposition 6.3.

Step V: The strong balancing property, (6.13) and (6.14) $\Rightarrow \mathbb{P}(A_1^c \cup A_2^c \cup \widetilde{B}_1^c \cup \widetilde{B}_2^c \cup B_3^c) \leq 5\gamma$.

We will start by bounding $\mathbb{P}(\widetilde{B}_1^c)$ and $\mathbb{P}(\widetilde{B}_2^c)$. Note that by Proposition 6.11 (ii) it follows that $\mathbb{P}(\widetilde{B}_1^c) \leq \gamma$ and $\mathbb{P}(\widetilde{B}_2^c) \leq \gamma$ as long as the strong balancing property is satisfied and

$$1 \gtrsim \Lambda \cdot \log\left(\frac{4}{\gamma}(\tilde{\theta} - s)\right), \qquad 1 \gtrsim \Upsilon \cdot \log\left(\frac{4}{\gamma}(\tilde{\theta} - s)\right) \tag{6.81}$$

where $\tilde{\theta} = \tilde{\theta}(\{q_k^i\}_{k=1}^r, 1/8, \{N_k\}_{k=1}^r, s, M)$ for $i = 1, 2$ and where $\tilde{\theta}$ is defined in Proposition 6.11 (ii) and Λ and Υ are defined in (6.64) and (6.65). Note that it is easy to see that we have

$$\left|\left\{ i \in \mathbb{N} : \max_{\substack{\Gamma_1 \subset \{1,\ldots,M\}, \quad |\Gamma_1| = s \\ \Gamma_{2,j} \subset \{N_{j-1}+1,\ldots,N_j\}, \quad j=1,\ldots,r}} \|P_{\Gamma_1} U^*((q_1^i)^{-1} P_{\Gamma_{2,1}} \oplus \ldots \oplus (q_r^i)^{-1} P_{\Gamma_{2,r}}) U e_i\| > \frac{1}{8\sqrt{s}} \right\}\right| \leq \tilde{M},$$

where

$$\tilde{M} = \min\{i \in \mathbb{N} : \|P_N U P_{i-1}^\perp\| \leq 1/(K32\sqrt{s})\},$$

and this follows from the choice in (6.52) where $q_k^1 = q_k^2 = \frac{1}{4} q_k$ for $1 \leq k \leq r$. Thus, it immediately follows that (6.13) and (6.14) imply (6.81).

As for bounding $\mathbb{P}(A_1^c)$, $\mathbb{P}(A_2^c)$ and $\mathbb{P}(B_3^c)$, this is done exactly as in Step V of the proof of Proposition 6.3.

Step VI: The strong balancing property, (6.13) and (6.14) $\Rightarrow \mathbb{P}(B_4^c) \leq \gamma$.

To see this, define the random variables $X_1, \ldots X_{u-2}$ as in (6.68). As in Step VI of the proof of Proposition 6.3 it suffices to show that (6.13) and (6.14) imply

$$\frac{1}{2} \geq \mathbb{P}(X_j = 1), \qquad j = 1, \ldots, u-2, \tag{6.82}$$

Claim: The strong balancing property, (6.13) and (6.14) $\Rightarrow$ **(6.82).** To prove the claim we first observe that $X_j = 0$ when

$$\|(P_\Delta - P_\Delta U^*(\frac{1}{q_1^i} P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i} P_{\Omega_r^i}) U P_\Delta) Z_{i-1}\|_{l^\infty} \leq \frac{1}{2}\|Z_{i-1}\|_{l^\infty}$$

$$\|P_\Delta^\perp U^*(\frac{1}{q_1^i} P_{\Omega_1^i} \oplus \ldots \oplus \frac{1}{q_r^i} P_{\Omega_r^i}) U P_\Delta Z_{i-1}\|_{l^\infty} \leq \frac{1}{4}\log_2(4K\tilde{M}\sqrt{s})\|Z_{i-1}\|_{l^\infty}, \qquad i = j+2.$$

Thus, by again recalling from (6.52) that $q_k^3 = q_k^4 = \ldots = q_k^u = \tilde{q}_k$, $\quad 1 \leq k \leq r$, and by choosing $\tilde{\gamma} = 1/4$ in (6.41) in Proposition 6.12 and $\tilde{\gamma} = 1/4$ in (ii) in Proposition 6.11, we conclude that (6.82) follows when the strong balancing property is satisfied as well as (6.72) and (6.73). and

$$\frac{\log_2(4K\tilde{M}\sqrt{s})}{\log\left(16(\tilde{M} - s)\right)} \geq C_2 \cdot \tilde{q}_k^{-1} \cdot (\nu_{k,1} + \ldots + \nu_{k,r}), \quad 1 \leq k \leq r \tag{6.83}$$

$$\frac{\log_2(4K\tilde{M}\sqrt{s})}{\log\left(16(\tilde{M} - s)\right)} \geq C_2 \cdot \left((\tilde{q}_1^{-1} - 1) \cdot \mu_{N_0} \cdot \tilde{s}_1 + \ldots + (\tilde{q}_r^{-1} - 1) \cdot \mu_{N_{r-1}} \cdot \tilde{s}_r\right), \tag{6.84}$$

for $K = \max_{1\le k\le r}(N_k - N_{k-1})/m_k$. for some constants C_1 and C_2. Thus, to prove the claim we must demonstrate that (6.13) and (6.14) $\Rightarrow$ (6.72), (6.73), (6.83) and (6.84). This is done by repeating Case 1 and Case 2 in Step VI of the proof of Proposition 6.3 almost verbatim, except replacing M by $\tilde{M}$. $\square$

6.4 Proof of Theorem 3.2

We now prove Theorem 3.2. First it is necessary to describe the wavelet construction in more detail. We refer to [7] for more details. Let $[0,a]$ be a compact interval and suppose that we are given an orthonormal mother wavelet Ψ and an orthonormal scaling function Φ such that $\mathrm{supp}(\Psi) = \mathrm{supp}(\Phi) = [0,a]$ for some $a \ge 1$. We also assume that for some $\alpha \ge 1$ and $C > 0$,

$$\left|\hat{\Phi}(\xi)\right| \le \frac{C}{(1+|\xi|)^{\alpha}}, \quad \left|\hat{\Psi}(\xi)\right| \le \frac{C}{(1+|\xi|)^{\alpha}}. \tag{6.85}$$

The most standard approach is to consider the following collection of functions

$$\Omega_a = \{\Phi_k, \Psi_{j,k} : \mathrm{supp}(\Phi_k)^o \cap [0,a] \neq \emptyset,\ \mathrm{supp}(\Psi_{j,k})^o \cap [0,a] \neq \emptyset, j \in \mathbb{Z}_+, k \in \mathbb{Z}, \},$$

where

$$\Phi_k = \Phi(\cdot - k), \qquad \Psi_{j,k} = 2^{\frac{j}{2}}\Psi(2^j \cdot -k).$$

(the notation K^o denotes the interior of a set $K \subseteq \mathbb{R}$). This now gives

$$\left\{f \in \mathrm{L}^2(\mathbb{R}) : \mathrm{supp}(f) \subseteq [0,a]\right\} \subseteq \overline{\mathrm{span}\{\varphi : \varphi \in \Omega_a\}} \subseteq \left\{f \in \mathrm{L}^2(\mathbb{R}) : \mathrm{supp}(f) \subseteq [T_1, T_2]\right\},$$

where $T_1, T_2 > 0$ are such that $[-T_1, T_2]$ contains the support of all functions in Ω_a. Note that the inclusions may be proper (but not always, as is the case with the Haar wavelet.) It is easy to see that

$$\Psi_{j,k} \notin \Omega_a \Longleftrightarrow \frac{a+k}{2^j} \le 0, \quad a \le \frac{k}{2^j},$$
$$\Phi_k \notin \Omega_a \Longleftrightarrow a + k \le 0, \quad a \le k,$$

and therefore

$$\Omega_a = \{\Phi_k : |k| = 0, \dots, \lceil a \rceil - 1\} \cup \{\Psi_{j,k} : j \in \mathbb{Z}_+, k \in \mathbb{Z}, -\lceil a \rceil < k < 2^j \lceil a \rceil\}.$$

We order Ω_a in increasing order of wavelet resolution as follows:

$$\begin{aligned}&\{\Phi_{-\lceil a\rceil+1}, \dots, \Phi_{-1}, \Phi_0, \Phi_1, \dots, \Phi_{\lceil a\rceil-1},\\ &\Psi_{0,-\lceil a\rceil+1}, \dots, \Psi_{0,-1}, \Psi_{0,0}, \Psi_{0,1}, \dots, \Psi_{0,\lceil a\rceil-1}, \Psi_{1,-\lceil a\rceil+1}, \dots\}.\end{aligned} \tag{6.86}$$

By the definition of Ω_a, we let $T_1 = \lceil a \rceil - 1$ and $T_2 = 2\lceil a \rceil - 1$.

Having constructed an orthonormal wavelet system form $[0,a]$ we now introduce the appropriate Fourier sampling basis. We must sample at at least the Nyquist rate. Hence we let $\epsilon \le 1/(T_1 + T_2)$ be the *sampling density* (note that $1/(T_1+T_2)$ is the Nyquist criterion for functions supported on $[-T_1, T_2]$) and define

$$\psi_j(x) = \sqrt{\epsilon}e^{2\pi i j \epsilon x}\chi_{[-T_1/(\epsilon(T_1+T_2)), T_2/(\epsilon(T_1+T_2))]}(x),$$

This gives an orthonormal sampling basis for $L^2[-T_1, T_2]$.

Proof of Theorem 3.2. Note that $\mu(U) \ge |\langle \Phi, \psi_0\rangle|^2 = \epsilon\left|\hat{\Phi}(0)\right|^2$, which gives the first result.

To show that $\mu(P_N^{\perp}U) = \mathcal{O}\left(N^{-1}\right)$, observe that the decay estimate in (6.85) yields

$$\begin{aligned}\mu(P_N^{\perp}U) &\le \max_{|k|\ge\frac{N}{2}}\max_{\varphi\in\Omega_a}|\langle\varphi,\psi_k\rangle|^2\\ &= \max\left\{\max_{|k|\ge\frac{N}{2}}\max_{R\in\mathbb{N}}\frac{\epsilon}{2^R}\left|\hat{\Psi}\left(\frac{-2\pi i\epsilon k}{2^R}\right)\right|^2, \max_{|k|\ge\frac{N}{2}}\left|\hat{\Phi}\left(-2\pi i\epsilon k\right)\right|^2\right\}\\ &\le \max_{|k|\ge\frac{N}{2}}\max_{R\in\mathbb{N}}\frac{\epsilon}{2^R}\frac{C^2}{\left(1+|2\pi\epsilon k 2^{-R}|\right)^{2\alpha}}\\ &\le \max_{R\in\mathbb{N}}\frac{\epsilon}{2^R}\frac{C^2}{\left(1+|\pi\epsilon N 2^{-R}|\right)^{2\alpha}}.\end{aligned}$$

The function $f(x) = x^{-1}(1 + \pi\epsilon N/x)^{-2\alpha}$ on $[1, \infty)$ is such that $f'(\pi\epsilon N(2\alpha - 1)) = 0$. Hence,

$$\mu(P_N^\perp U) \le \frac{C^2}{\pi N(2\alpha - 1)(1 + 1/(2\alpha - 1))^{2\alpha}}$$

which gives $\mu(P_N^\perp U) = \mathcal{O}\left(N^{-1}\right)$.

Let $\Omega_{R,a}$ contain all wavelets in Ω_a with resolution less than R, so

$$\Omega_{R,a} = \{\varphi \in \Omega_a : \varphi = \Psi_{j,k},\ j < R,\ k \in \mathbb{Z} \text{ or } \varphi = \Phi_k,\ k \in \mathbb{Z}\}.$$

Then, denoting the size of $\Omega_{R,a}$ by $N^{(R)}$, it is easy to verify that $N^{(R)} = 2^R \lceil a \rceil + (R+1)(\lceil a \rceil - 1)$. Given any $N \in \mathbb{N}$ such that $N \ge N^{(1)}$, let R be such that

$$N^{(R)} \le N < N^{(R+1)}.$$

Then, for each $n \ge N$, there exists some $j \ge R$ and $l \in \mathbb{Z}$ such that the n^{th} element via the ordering (6.86) is $\varphi_n = \Psi_{j,l}$.

$$\begin{aligned}
\mu(UP_N^\perp) &= \max_{n \ge N} \max_{k \in \mathbb{Z}} \left|\langle \varphi_n, \psi_k \rangle\right|^2 \\
&= \max_{j \ge R} \max_{k \in \mathbb{Z}} \frac{\epsilon}{2^j} \left| \hat{\Psi}\left(\frac{-2\pi i \epsilon k}{2^j} \right) \right|^2 \\
&\le \|\hat{\Psi}\|_{L^\infty}^2 \frac{\epsilon}{2^R} \\
&\le 4 \|\hat{\Psi}\|_{L^\infty}^2 \frac{\epsilon \lceil a \rceil}{N},
\end{aligned}$$

where the last line follows because $N < N^{(R+1)} = 2^{R+1}\lceil a \rceil + (R+2)(\lceil a \rceil - 1)$ implies that

$$2^{-R} < \frac{1}{N}\left(2\lceil a \rceil + (R+2)(\lceil a \rceil - 1)2^{-R}\right) \le \frac{4\lceil a \rceil}{N}.$$

So $\mu(UP_N^\perp) = \mathcal{O}\left(N^{-1}\right)$. □

7 Numerical examples

In this section we present the numerical examples to support our theory.

7.1 The GLPU Phantom

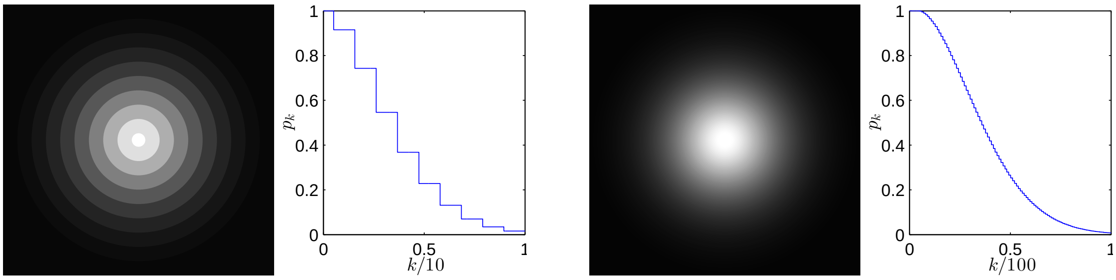
To illustrate the main message that asymptotic sparsity and asymptotic incoherence yield the phenomenon that the success of compressed sensing is resolution dependent, it is preferable to work with a continuous model, which is closer to real life MRI scenarios. This is impossible if one employs an already discretized model at a fixed resolution level. For this reason we choose to use the novel GLPU phantom invented by Guerquin–Kern, Lejeune, Pruessmann and Unser in [30] in favour of the standard discretized Shepp-Logan Phantom from MATLAB. The GLPU phantom is a so-called analytic phantom, in that it is not a rasterized image, but rather a continuous (or *infinite-resolution*) object defined by analytic curves, such as Bezier curves. The MATLAB code offered by the authors allows one to compute the continuous (integral) Fourier transform (as in a real life MRI scenario) of the GLPU phantom, for any resolution, to avoid the inverse crime which results from using the discrete Fourier transform.

We thus treat the GLPU phantom as a function $f : D \to \mathbb{R}$ which is measured by taking equispaced pointwise samples of its continuous Fourier transform:

$$\hat{f}(\omega) = \int_D f(x) \mathrm{e}^{-2\pi \mathrm{i} \omega \cdot x} \, \mathrm{d}x,$$

where D is typically $[0, 1]^2$.

We note that full resolution and uncropped versions of the images used in this section are available online [1].



(a) $n{=}10$, $m{=}0.05$, $a{=}1.75$, $b{=}2.25$ (b) $n{=}100$, $m{=}0.05$, $a{=}1.75$, $b{=}2.4726$

Figure 6: Examples of subsampling maps at 2048×2048 that subsample $p = 15\%$ of Fourier coefficients. Left: 10 levels, right: 100 levels. The color intensity denotes the fraction p_k of random samples taken uniformly, i.e. white: 100% samples, black: 0% samples.

7.2 Subsampling scheme

Our subsampling scheme divides the 2D Fourier spectrum of $N \times N$ coefficients into n regions delimited by $n-1$ equispaced concentric circles plus the full square, an example being shown in Figure 6. Normalizing the 2D Fourier spectrum to $[-1,1]^2$, the circles have radius r_k with $k = 0, \dots, n-1$, which are given by $r_0 = m$ and $r_k = k \cdot \frac{1-m}{n-1}$ for $k > 0$, where $0 \leq m < 1$ is a parameter. Inside each of the n regions, the fraction p_k of Fourier coefficients taken with uniform probability is given by:

$$p_k = \exp\left(-\left(b \cdot \frac{k}{n}\right)^a\right), \tag{7.1}$$

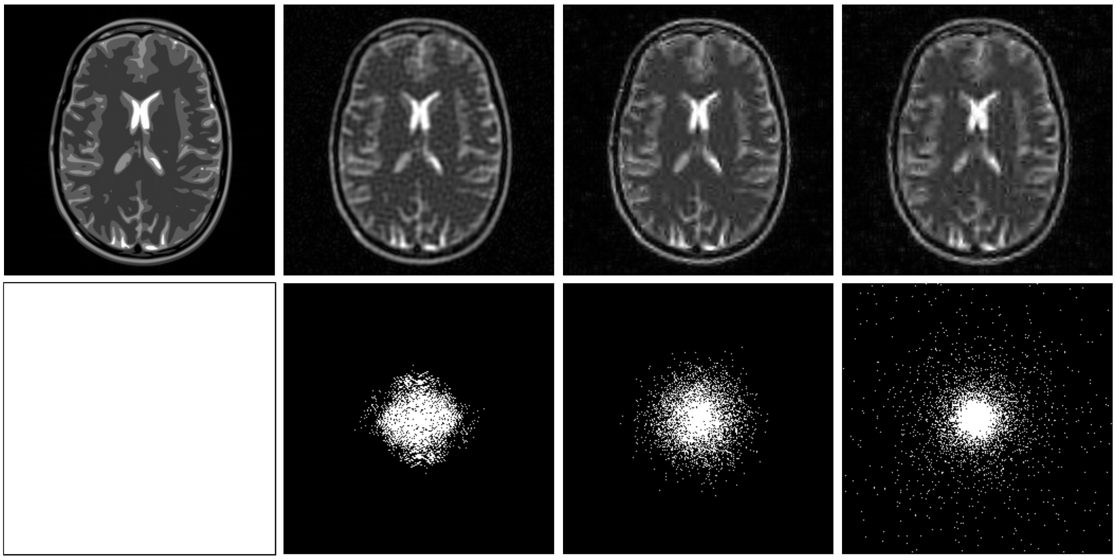
where $k = 0, \dots, n$ and $a > 0$ and $b > 0$ are parameters, so the total fraction of coefficients that are subsampled from the full spectrum is then $p = \sum_k p_k S_k$, where S_k is the normalized area of the kth region. It is obvious that the first region will sample all Fourier coefficients ($p_0 = 1$) and the remaining regions will sample a monotonically smaller fraction of coefficients ($p_{k+1} < p_k$). The function (7.1) is very similar to the generalized Gaussian distribution function. Although we obtained good results with this subsampling scheme, we do not claim that it is optimal, nor the best *all-rounder*. As stated and shown previously, an optimal subsampling scheme is highly dependent on the signal structure (i.e. image content in this case) and also resolution dependent. Throughout this section we use $n = 100$ levels, but (7.1) is versatile enough and we obtained similar reconstruction errors with smaller values of n.

The reasons we used such a subsampling scheme include a high degree of control for its shape, as well as a good match with the behaviour of the Fourier spectrum and the sparsity of real-life images, e.g. MRI images. Knowing that for a fixed M, the first $M \times M$ Fourier coefficients do not change when the image resolution N increases (i.e. N grows), and that the image is asymptotically sparse and asymptotically incoherent in the chosen wavelet basis as resolution increases, it is thus desirable that the subsampling scheme reshapes as a function of the resolution to achieve the lowest error when a constant fraction p of coefficients is to be subsampled. Using a normalized spectrum for (7.1) achieves that purpose to an extent[2], while the parameters a, b can be made as functions of N and of p so that the asymptotic sparsity is even better exploited as image resolution increases.

7.3 Resolution dependence: Fixed fraction p of samples

The important message here is that *regardless of the subsampling scheme used*, the quality of the reconstruction will increase as the resolution increases when compared to the full sampled version at the respective resolution. As previously explained, this is because at high resolution levels the image signal is increasingly sparse and incoherent, which can be fruitfully exploited.

[2] In this sense, (7.1) will assign different probabilities for the same 2D Fourier coefficient ω_1, ω_2 as the resolution N changes, since the n regions in (7.1) are defined on the normalized 2D Fourier spectrum.



(a) Full sampling. Relative error is 0%.

(b) Largest 5% of coefficients. Linear reconstruction. Relative error is 17.62%.

(c) Multi level, 5% of coefficients, Gaussian law (7.1). Relative error is 18.59%.

(d) 5% of coefficients, power law[3]. Relative error is 22.22%.

Figure 7: Subsampling 5% of Fourier coefficients. The bottom row shows the subsampling map used. The relative error to the full-sampled case is large regardless of the subsampling scheme used. The relative error is computed as $\|f - f_p\|_2/\|f\|_2$, where f and f_p represent the reconstructed full sampled image and subsampled image respectively.

At low resolution levels however, the sparsity level is also low and subsampling a small fraction p of Fourier coefficients is bound to give a poor quality reconstruction. In Figure 7 we show an experiment where we subsample 5% of Fourier coefficients from the first 256×256 using three different schemes and reconstruct using Daubechies 4 wavelets. As expected, the result is quite poor. We confirm that we tried a large number of different random subsampling schemes and wavelet bases and they all gave similarly poor results, visually and numerically. This may mislead one to believe that compressed sensing fails with high compression ration, i.e. low p.

A much different conclusion is reached when the experiment is repeated at higher resolutions. A first example that illustrate the resolution dependence of compressed sensing reconstruction is shown in Figure 8. In this experiment we fixed the parameters m, a, b of the Gaussian subsampling scheme 7.1, and subsampled $p = 5\%$ of Fourier coefficients at increasing image resolutions from 256×256 to 4096×4096, reconstructing in the Daubechies 4 wavelet basis. The high and asymptotic sparsity of the wavelet coefficients at high resolutions allows a markedly better quality reconstruction than at low resolutions when compared to the full sampled version of the same resolution.

We want to stress the fact that *any subsampling scheme* yields the same effect of resolution dependency, which was confirmed in our tests.

A further example is shown in Figure 9 where we sampled 10% of Fourier coefficients, again keeping $n = 100$ levels but optimizing a and b to give the best reconstruction in both cases. In this case, the high resolution 4096×4096 reconstruction has hardly any visible artefacts left. A fraction of $p = 10\%$ means an effective factor of 10 speed-up for an MRI, and at this resolution it is equivalent in weight to a full sampled image of 1296×1296, but with the tremendous advantage that it is actually of better quality and with more details than the full sampled 1296×1296 equivalent.

[3]The power law subsampling scheme involves sampling the 2D Fourier coefficient ω_1, ω_2 with probability $(\omega_1^2 + \omega_2^2 + 1)^{-\alpha}$, where $-N/2 < \omega_1, \omega_2 < N/2 - 1$ and α is a parameter here chosen as $\alpha = 3/2$.

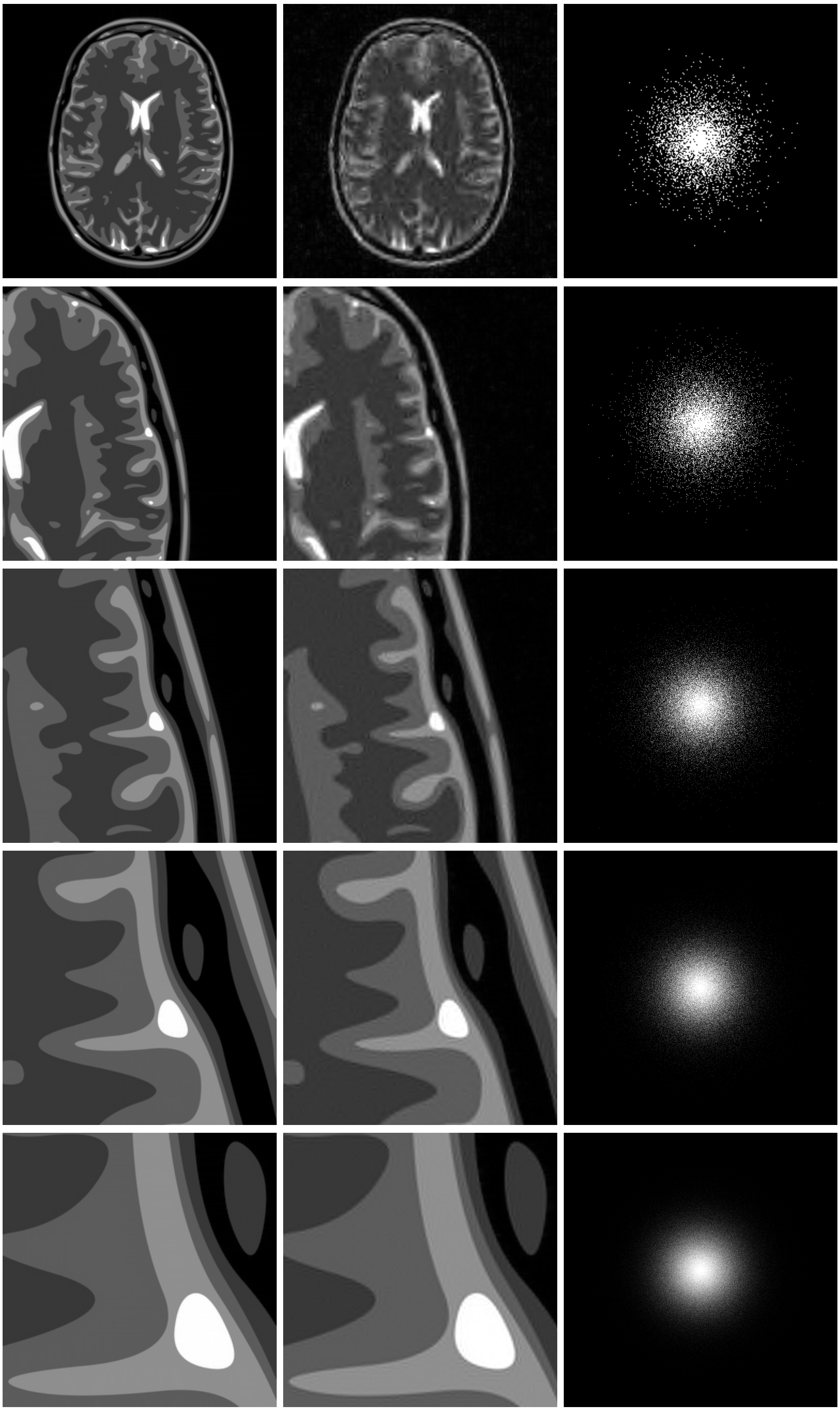
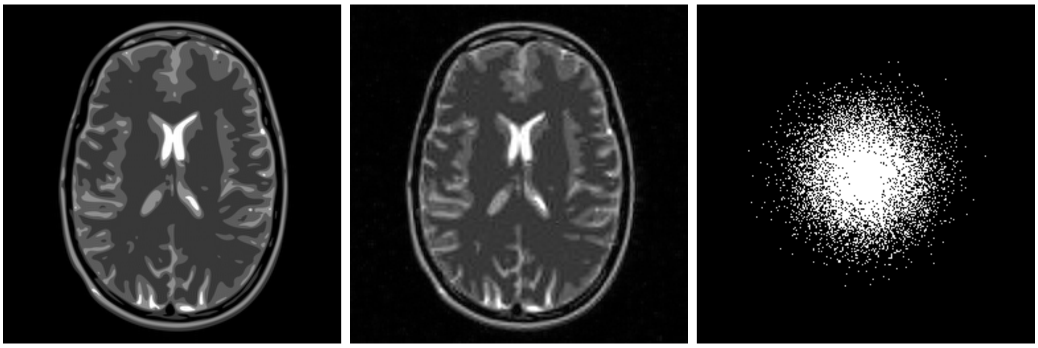


Figure 8: Multi-level subsampling of 5% Fourier coefficients using (7.1) with fixed parameters $n = 100$, $a = 1.75$, $b = 4.125$. The left column (full sampled) and center column (subsampled) are crops of 256×256 pixels of the original full resolution versions, while the right column shows the uncropped subsampling map used. The error shown is the relative error between the subsampled and full sampled versions.



(a) 256×256 full sampled (left) and 10% subsampled (center). Relative error to full sampling is 12.12%. Artefacts are obvious.

(b) 4096×4096 full sampled (left) and 10% subsampled (center), showing crops of 256×256 to preserve pixel size. Relative error to full sampling is 1.48%. Artefacts are hardly visible.

Figure 9: Improvement at 10% subsampling between resolutions. The subsampling map is shown in the right column. In (7.1)) we used $n = 100$ and $m = 0.01$ and varied a and b to obtain the best result in each case.

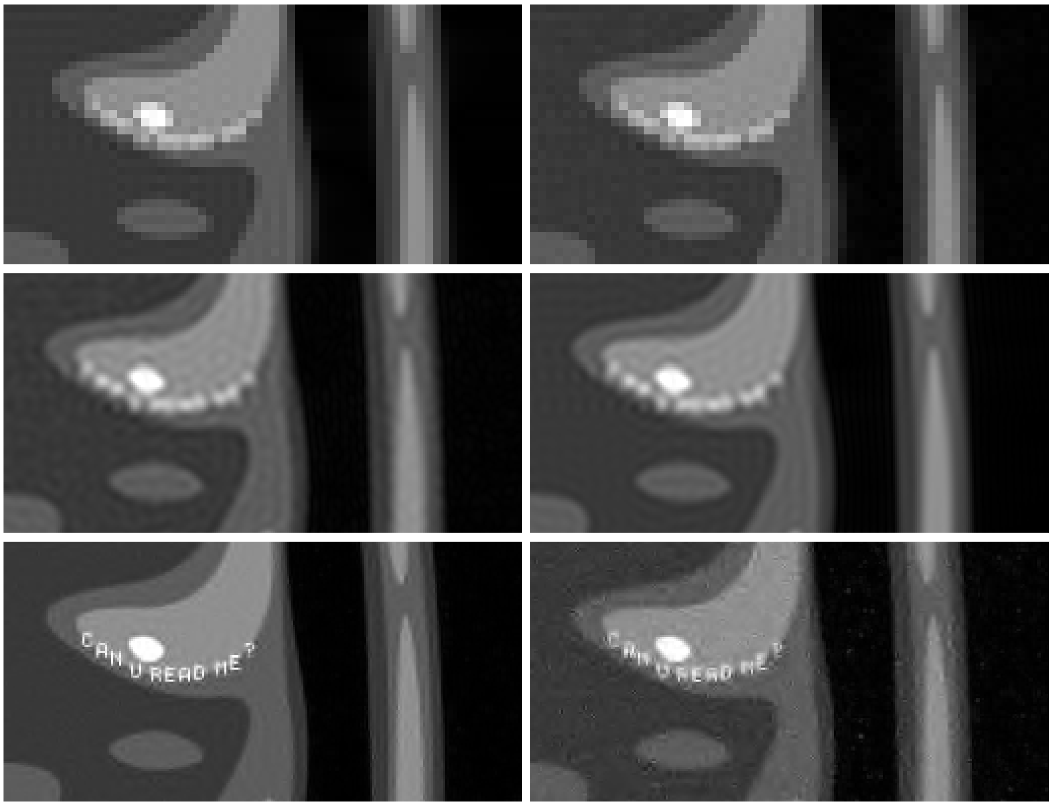
7.4 Resolution dependence: Fixed number of samples

The above result of resolution dependence with a fixed fraction p is due to the asymptotic sparsity and asymptotic incoherence, but is in part also due to the fact that a fixed fraction p does mean more coefficients being sampled as the resolution increases.

A more spectacular result of asymptotic sparsity and asymptotic incoherence is obtained by running a similar experiment, but this time fixing the number of coefficients being sampled, rather than the fraction p. This was done in Figure 10, where the same number of coefficients was sampled: $512^2 = 262144$ Fourier coefficients. Fine details were hidden in the image and then three reconstructions were performed: (a) the full sampled version of 512×512 pixels, (b) the linear reconstruction of the subsampled 2048×2048 version by zero-padding the first 512×512 coefficients, and (c) the mulitlevel subsampled 2048×2048 reconstruction using (7.1) and the Daubechies 4 wavelet basis.

Even though the number of coefficients was the same in all reconstruction cases, the higher resolution and multilevel subsampling means that the asymptotic sparsity and incoherence of wavelet coefficients can be fruitfully exploited, and the fine details recovered to a much clearer extent, even in the presence of noise.

This effectively means that by simply going higher in resolution or, equivalently, higher in the Fourier spectrum, one can recover a signal much closer to the exact one, yet taking the same amount of measurements.



(a)

(b)

(c)

Figure 10: Subsampling a fixed number of $512^2 = 262144$ Fourier coefficients, in three reconstruction scenarios. Left column: no noise. Right column: white noise, SNR = 16 dB. (a) 512×512 full sampled reconstruction. (b) 2048×2048 linear reconstruction from the first $512\times512 = 262144$ Fourier coefficients (zero padded). (c) 2048×2048 reconstructed from 262144 Fourier coefficients taken with a multi level scheme using (7.1) and Daubechies 4, with $m = 0.05$, $a = 1.25$, $b = 4.2539$.

8 Conclusions

The purpose of this paper was to bridge an important gap between compressed sensing theory and its use in imaging applications by introducing a framework based on asymptotic incoherence, asymptotic sparsity and multilevel sampling. In doing so, we have not only given mathematical credence to the abundance on numerical evidence suggesting the usefulness of variable density sampling strategies, but also drawn important conclusions about the resolution dependence of compressed sensing and the signal structure dependence of optimal sampling strategies.

Recently, Krahmer & Ward have shown recoverability in the discrete bivariate Haar wavelet basis from Fourier measurements taken according to an inverse square law [31]. Their analysis is valid for bivariate Haar wavelets with Fourier samples in the finite-dimensional setting for only one particular variable density. It is also based on the RIP, which, as discussed in Section 5.4, is problematic in imaging and does not take into account asymptotic sparsity, which is a key in real-life situations.

Our framework, on the other hand, allows for arbitrary sampling and sparsity systems in both the finite- and infinite-dimensional setting, as well as rather general types of sampling strategies, and avoids the RIP. We also make clear the interplay between asymptotic sparsity and asymptotic incoherence, and how they lead to the key conclusion of resolution dependence. Having said this, recoverability results for TV-minimization—an important and powerful technique in imaging—are also given in [31]. Whilst beyond the scope of this paper, we expect that our results can also be extended to the TV case. This is an objective of ongoing work.

As explained in the paper, traditional sparsity is not sufficient to describe the behaviour of

natural images. Asymptotic sparsity in levels more accurately describes natural images, and hence the behaviour of multilevel sampling strategies. We remark that we are by no means the first to advocate a more structured notion of sparsity. So-called *model-based* compressed sensing see [8, 24] and references therein) also move beyond sparsity. However, this is quite different to our approach. Algorithms such as those found in [8] seek to incorporate substantially more intricate models than standard sparsity (or, indeed, asymptotic sparsity in levels) so as to further reduce the number of measurements required in recovery problems that are already incoherent beyond that needed for standard compressed sensing algorithms. The main problem in [8] is that of sampling with random Gaussians. Although highly incoherent, this is not a suitable model MRI, where the measurement system (i.e. Fourier samples) is constrained by the physical device. On the other hand, the motivation in our work stems primarily from the desire to tackle the fixed, and high, coherence. The fact that the resulting algorithm of multilevel sampling is well suited for precisely the types of signals and images one sees in application is, in many senses, a serendipity.

We have concluded in our work that the optimal sampling strategy is dependent on the signal structure. Further work is required to determine such strategies in a rigorous empirical manner for important classes of images. We expect our main theorems to give important insights into these investigations, such as how many levels to choose, how to choose their relative sizes, etc. One conclusion of our work, however, is that approaches to design optimal sampling densities based solely on minimizing coherences (i.e. not taking into account asymptotic sparsity) may be of little use in practice unless they are trained on large families of images having similar structures (e.g. brain images).

Another current investigation involves deriving refined versions of our theorems (which are completely general) in the case of Fourier sampling with wavelets. In particular, we expect it to be possible to demonstrate that the 'interference' between sparsity levels (see Section 4.2) cannot be too bad in this setting, meaning that the key bounds on the number of measurements can be simplified to something similar to the block diagonal case (see Section 4.2.2). This work will build on the techniques developed in [7].

Finally, we would like to point out that asymptotic sparsity is not only relevant for wavelets. Any approximation system whose power lies in nonlinear, as opposed to linear, approximation will give rise to asymptotically sparse representations. Such systems include curvelets [9, 11], contourlets [20, 37] and shearlets [17, 18, 32], to name but a few. Hence we expect our framework to have relevance in numerous other applications.

Acknowledgements

The authors would like to thank Akram Aldroubi, Emmanuel Candès, Massimo Fornasier, Felix Krahmer, Thomas Strohmer, Gerd Teschke and Rachel Ward for useful discussions and comments.

References

[1] `http://subsample.org`, Feb. 2013.

[2] B. Adcock and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Technical report NA2011/02, DAMTP, University of Cambridge*, 2011.

[3] B. Adcock and A. C. Hansen. A generalized sampling theorem for stable reconstructions in arbitrary bases. *J. Fourier Anal. Appl.*, 18(4):685–716, 2012.

[4] B. Adcock and A. C. Hansen. Stable reconstructions in Hilbert spaces and the resolution of the Gibbs phenomenon. *Appl. Comput. Harmon. Anal.*, 32(3):357–388, 2012.

[5] B. Adcock, A. C. Hansen, E. Herrholz, and G. Teschke. Generalized sampling: extension to frames and inverse and ill-posed problems. *Inverse Problems*, (to appear).

[6] B. Adcock, A. C. Hansen, and C. Poon. Beyond consistent reconstructions: optimality and sharp bounds for generalized sampling, and application to the uniform resampling problem. *Preprint*, 2012.

[7] B. Adcock, A. C. Hansen, and C. Poon. On optimal wavelet reconstructions from Fourier samples: linearity and universality of the stable sampling rate. *Technical report NA2012/07, DAMTP, University of Cambridge*, 2012.

[8] R. G. Baraniuk, V. Cebher, M. F. Duarte, and C. Hedge. Model-based compressive sensing. *IEEE Trans. Inform. Theory*, 56(4):1982–2001, 2010.

[9] E. Candès and D. L. Donoho. Recovering edges in ill-posed inverse problems: optimality of curvelet frames. *Ann. Statist.*, 30(3):784–842, 2002.

[10] E. J. Candès. An introduction to compressive sensing. *IEEE Signal Process. Mag.*, 25(2):21–30, 2008.

[11] E. J. Candès and D. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.

[12] E. J. Candès and Y. Plan. A probabilistic and RIPless theory of compressed sensing. *IEEE Trans. Inform. Theory*, 57(11):7235–7254, 2011.

[13] E. J. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007.

[14] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[15] Y. Chi, L. L. Scharf, A. Pezeshki, and A. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Trans. Signal Process.*, 59(5):2182–2195, 2011.

[16] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. Signal Process.*, 46:886–902, 1998.

[17] S. Dahlke, G. Kutyniok, P. Maass, C. Sagiv, H.-G. Stark, and G. Teschke. The uncertainty principle associated with the continuous shearlet transform. *Int. J. Wavelets Multiresolut. Inf. Process.*, 6(2):157–181, 2008.

[18] S. Dahlke, G. Kutyniok, G. Steidl, and G. Teschke. Shearlet coorbit spaces and associated Banach frames. *Appl. Comput. Harmon. Anal.*, 27(2):195–214, 2009.

[19] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to compressed sensing. In *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2011.

[20] M. N. Do and M. Vetterli. The contourlet transform: An efficient directional multiresolution image representation. *IEEE Trans. Image Proc.*, 14(12):2091–2106, 2005.

[21] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[22] D. L. Donoho and M. Elad. Optimally sparse representation in general (non-orthogonal) dictionaries vi ℓ_1 minimizatiob. *Proc. Natl Acad. Sci. USA*, 100:2197–2002, 2003.

[23] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47:2845–2862, 2001.

[24] M. F. Duarte and Y. C. Eldar. Structured compressed sensing: from theory to applications. *IEEE Trans. Signal Process.*, 59(9):4053–4085, 2011.

[25] Y. C. Eldar and G. Kutyniok, editors. *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[26] M. Fornasier and H. Rauhut. Compressive sensing. In *Handbook of Mathematical Methods in Imaging*, pages 187–228. Springer, 2011.

[27] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. In preparation, 2013.

[28] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Theor.*, 57(3):1548–1566, Mar. 2011.

[29] M. Guerquin-Kern, M. Häberlin, K. Pruessmann, and M. Unser. A fast wavelet-based reconstruction method for magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 30(9):1649–1660, 2011.

[30] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser. Realistic analytical phantoms for parallel Magnetic Resonance Imaging. *IEEE Trans. Med. Imaging*, 31(3):626–636, 2012.

[31] F. Krahmer and R. Ward. Compressive imaging: stable and robust recovery from variable density frequency samples. *Preprint*, 2012.

[32] G. Kutyniok, J. Lemvig, and W.-Q. Lim. Compactly supported shearlets. In M. Neamtu and L. Schumaker, editors, *Approximation Theory XIII: San Antonio 2010*, volume 13 of *Springer Proceedings in Mathematics*, pages 163–186. Springer New York, 2012.

[33] M. Lustig. *Sparse MRI*. PhD thesis, Stanford University, 2008.

[34] M. Lustig, D. L. Donoho, and J. M. Pauly. Sparse MRI: the application of compressed sensing for rapid MRI imaging. *Magn. Reson. Imaging*, 58(6):1182–1195, 2007.

[35] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed Sensing MRI. *IEEE Signal Process. Mag.*, 25(2):72–82, March 2008.

[36] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.

[37] D. D.-Y. Po and M. N. Do. Directional multiscale modeling of images using the contourlet transform. *IEEE Trans. Image Proc.*, 15(6):1610–1620, June 2006.

[38] G. Puy, J. P. Marques, R. Gruetter, J. Thiran, D. Van De Ville, P. Vandergheynst, and Y. Wiaux. Spread spectrum Magnetic Resonance Imaging. *IEEE Trans. Med. Imaging*, 31(3):586–598, 2012.

[39] G. Puy, P. Vandergheynst, and Y. Wiaux. On variable density compressive sampling. *IEEE Signal Process. Letters*, 18:595–598, 2011.

[40] M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164(1):60–72, 1999.

[41] T. Strohmer. Measure what should be measured: progress and challenges in compressive sensing. *IEEE Signal Process. Letters*, 19(12):887–893, 2012.

[42] V. Studer, J. Bobin, M. Chahid, H. Moussavi, E. Candès, and M. Dahan. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Submitted*, 2011.

[43] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.

[44] J. A. Tropp. On the conditioning of random subdictionaries. *Appl. Comput. Harmon. Anal.*, 25(1):1–24, 2008.

[45] Z. Wang and G. R. Arce. Variable density compressed image sampling. *IEEE Trans. Image Proc.*, 19(1):264–270, 2010.